Structural studies on C3 exoenzymes and botulinum neurotoxins

Evans, Hazel

Award date:
2003

Awarding institution:
University of Bath

Structural studies on C3 exoenzymes and Botulinum neurotoxins

Submitted by Hazel Evans for the degree of PhD of the University of Bath

Copyright

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

For the glory of God, the most high, the almighty one.

'The vision is yet for the appointed time; and it hasteth towards the end and shall not lie. Though it tarries, wait for it, because it will surely come.'

Habakkuk 2v3, The Bible (KJV)

Thanks to...

David Lawton for all the cups of tea.
Simon, Alison and Andrew for all the prayers
My parents for all their prayers

The staff at the SRS, Daresbury

All the people who helped me in the lab and with my computer: Natesh, Shalini, Matt, Michelle, Jawahar and particularly Gayatri for teaching me lots of stuff

Dan Holloway
For all the help in the lab and attention to detail.

Joanne Ayriss
For helping with the cells and the Biacore, not to mention the lifts and the coffee breaks.

Mark Sutton
For the clones, the expression, help with everything else and the wacky ideas.

Ravi Acharya and Cliff Shone
For all their help, advice and support.

The University of Bath and the Health Protection Agency, Porton Down for funding.

The Lord Jesus Christ, the creator of the world who sustains all things by his powerful word

Abbreviations

ADP	Adenosine diphosphate
ADPRTs	ADP-ribosylating toxins
ARTT	ADP-ribosylating turn turn
βOG	octyl-β-D-glucopyranoside
B. cereus	*Bacillus cereus*
BoNTs	Botulinum neurotoxins
BSA	Bovine serum albumin
C3	C3bot1 from *Clostridium botulinum*
C. botulinum	*Clostridium botulinum*
CCD	Charge-coupled device
CHAPS	3-[-(cholamidopropyl)dimethylammoniol]-1-propanesulfonate
C. limosum	*Clostridium limosum*
CNS	Crystallographic and NMR System
CT	Cholera toxin
Cu	Copper
DNA	Deoxyribonucleic acid
DMEM	Dulbecco Modified Eagle Medium
dNTPs	deoxynucleotide triphosphates
DT	Diphtheria toxin
DTT	Dithiothreitol
E. coli	*Escherichia coli*
EDIN	Epidermal Differentiation Inhibitor
EDTA	Ethylenediaminetetraacetic acid
ETA	*Pseudomonas aeruginosa* Exotoxin A
GAP	GTPase activating protein
GDI	GDP dissociation inhibitors
GDP	Guanine diphosphate
GEF	Guanine exchange factor
GST	Glutathione-S-transferase
GTP	Guanine triphosphate
gui	graphical user interface
HC	Heavy chain (Binding and translocation domain) of BoNTs

H_C	C-terminus (binding domain) of heavy chain of BoNTs
HPA	Health Protection Agency, Porton Down
KJV	King James Version
MBP	Maltose binding protein
$MgCl_2$	Magnesium chloride
$MgSO_4$	Magnesium sulphate
MIR	Multiple Isomorphous Replacement
ML	Maximum Likelihood
MPD	2-methly-2,4-pentanediol
Na	Sodium
NaCl	Sodium Chloride
NAD	Nicotinamide adenine dinucleotide
NADH	Nicotinamide adenine dinucleotide (reduced)
Na_2KPO_4	Sodium potassium phosphate
NaOH	Sodium hydroxide
$(NH_4)_2SO_4$	Ammonium sulphate
NMA	5-n-acetyl-alpha-d-neuraminic acid
NMR	Nuclear Magnetic Resonance
PCR	Polymerase chain reaction
PDB/pdb	Protein data bank. Used to describe coordinate files. In figure legends it refers to the PDB code from the protein data bank.
PEG	Polyethylene glycol
PMSF	Phenlymethylsulphonyl fluoride
PN	Phosphate -nicotinamide
r.m.s.	Root mean square
S. aureus	*Staphylococcus aureus*
SDS-PAGE	Sodium dodecyl sulphate -Polyacrylamide Gel Electrophoresis
SNAP	Soluble NSF attachment protein
SNARE	SNAP receptor
SPR	Surface plasmon resonance
SRS	Synchrotron Radiation Source
TAE	Tri-acetate/EDTA electrophoresis buffer
TCA	Trichloroacetic acid
VIP	Vegetative Insecticidal protein from *Bacillus Cereus*

Materials

All chemicals, mentioned in this thesis were obtained from Sigma unless otherwise stated.

Abstract

One of the major fascinations with protein biochemistry is the way that proteins can discriminate between similar substrates to perform specific reactions. This specificity is very evident in the cell where communication relies on these interactions. Even more incredible, however, is the ability of bacteria to evolve toxins which are also highly specific for particular eukaryotic cellular targets, in order to elicit particular cell responses. The C3 exoenzymes specifically ADP-ribosylate the Rho GTPases RhoA-C, and sometimes RhoE/Rnd3, but do not ADP-ribosylate the 55% identical GTPases Rac and Cdc42. The botulinum neurotoxins are delivered specifically to the neuromuscular junctions via their cell binding domains rather than to the inhibitory neurones of the spinal cord as for the homologous tetanus neurotoxin. In addition, there is great medical interest in engineering toxins such as these for therapeutics. This thesis presents research to forward the understanding of these molecules and their specificity and to aid further interest in engineering them for therapeutic use.

As this thesis contains structural data from X-ray crystallography, chapter 1 outlines its general principles and gives an overview of the practice of this technique. Chapter 2 provides a general structural background to ADP-ribosylating toxins and Rho GTPases, a structural and mechanistic overview of the botulinum neurotoxins as well as an in depth literature review of the C3 exoenzyme from *Clostridium botulinum*. The particular context of each piece of work is presented in the individual chapter introductions. Chapter 3 presents the purification, crystallisation and structure determination of C3 (C3bot1) from *Clostridium botulinum* in a previously undetermined crystal form. Chapter 4 presents the expression, purification, crystallisation and structure determination of C3stau2 from *S. aureus* and its complex with NAD. It compares NAD binding between the C3-like and binary toxins and comments on the difference in protein target specificity between C3bot1 and C3stau2. Chapter 5 presents a basic functional characterisation of C3bot1 and C3stau2 as well as preliminary experiments to investigate C3-Rho binding. Chapter 6 describes the attempts to crystallise the H_C domain of botulinum neurotoxins A and F. A brief final section comments on the status of, and future for, research in these areas.

Contents

Chapter 1: Protein Crystallography

1.1	Introduction	2
1.2	Crystals and diffraction	2
1.3	Protein expression and purification	7
1.4	Crystallisation	9
1.5	Crystal manipulation	11
1.6	X-ray Sources	12
1.7	Data collection	12
1.8	Data processing and scaling	14
1.9	Phasing methods	15
	1.9.1 Novel phasing methods	16
	1.9.2 Molecular replacement	17
1.10	Electron density maps, density modification and model building	19
1.11	Refinement	21
1.12	Structural validation	23

Chapter 2: ADP-ribosylating toxins and Botulinum neurotoxins

2.1	Introduction	25
2.2	ADP-ribosylating toxins	25
2.3	Rho GTPase structure and regulation	32
2.4	A history of C3 from *Clostridium botulinum*	41
2.5	Botulinum neurotoxin structure and mechanism	48

Chapter 3: The Structure of C3bot1

3.1	Introduction	54
3.2	Methods and Results	54
	3.2.1 Purification of C3bot1	54

3.2.2 Crystallisation 57
3.2.3 Data collection 58
3.2.4 Refinement 59
3.2.5 Structure and crystal packing of C3bot1 63
3.3 Discussion 69

Chapter 4: The Structure of C3stau2 and its Complex with NAD

4.1 Introduction 73
4.2 Methods and Results 75
4.2.1 Cloning and expression 75
4.2.2 Protein purification 79
4.2.3 Crystallisation 79
4.2.4 Data collection and processing for native C3stau2 80
4.2.5 Refinement and model building for native C3stau2 82
4.2.6 Data Collection and Processing for the C3stau2-NAD complex 84
4.2.7 Refinement and model building for the C3stau2-NAD complex 85
4.3 Results and Discussion 88
4.3.1 Overall topology of C3stau2 88
4.3.2 The ARTT loop 89
4.3.3 The PN binding loop 93
4.3.4 C3stau2 moves to enclose NAD 93
4.3.5 Analysis of the common NAD-binding site of the C3-like and binary toxins 95
4.3.6 Implications for RhoA and RalA binding 97
4.3.7 Further Thoughts 100

Chapter 5: Characterisation of C3bot1 and C3stau2

5.1 Introduction 104
5.2 Methods 106
5.3 Results and Discussions 110
5.3.1 Purification of RhoA and RalA 110

a, Results 110
b, Discussion 112
5.3.2 Characterisation of C3bot1 113
a, ADP-ribosyltransferase activity 113
b, NAD glycohydrolase activity 113
c, Cell assays 115
d, Discussion 117
5.3.3 Characterisation of C3stau2 119
a, ADP-ribosyltransferase activity 119
b, NAD glycohydrolase activity 120
c, Cell assays 120
d, Discussion 120
5.3.4 Investigation of the C3bot1-RhoA complex 123
a, NAD analogue experiments 123
b, GST binding assays 126
c, Gel filtration binding assays 127
d, Surface plasmon resonance experiments 129
e, Discussion 131
5.3.5 Investigation of the C3bot1-RalA complex 132
a, Gel filtration binding assays 132
b, Surface plasmon resonance experiments 132
c, Discussion 134

Chapter 6: Crystallisation of Botulinum neurotoxins Hc/A and Hc/F

6.1 Introduction 136
6.2 Methods 137
6.3 Results 138
6.4 Discussion 142

Conclusions and further work 143

References 148

Appendix- Crystallographic programs 165

Chapter 1

Protein Crystallography

1.1 Introduction

The technique of x-ray crystallography is one of many techniques which can be used to study the structures of biological macromolecules in order to understand how they function. Amongst these techniques, however, crystallography is unique in its ability to provide a detailed, high resolution picture of molecular structures. In essence, x-rays are used to take a 'photo' of the molecule. The need for x-rays is due to the details of molecular structure being smaller than the wavelength of visible light which is between 400-700nm. Covalently bonded atoms therefore require x-ray radiation with its wavelength between 0.1-1nm to resolve them. The use of x-rays to visualise molecules adds the complication that x-rays diffracted from an object cannot be focused back onto a screen. Therefore, the unfocussed information must be collected and converted back to an image using a computational process. A further complication is that single molecules do not diffract x-rays strongly, most x-rays will pass straight through. To overcome this obstacle, crystals containing repeating arrays of the molecule are grown to produce a larger signal. This chapter will discuss the steps involved in protein crystallography from producing protein crystals through to structure determination in order to study protein function.

1.2 Crystals and Diffraction

Crystals comprise a lattice of a regularly repeated unit known as the unit cell. The unit cell is defined as the smallest building block for a crystal which can be built into the entire crystal using only translation functions. Unit cells are defined by the length of their three axis, a, b and c, and the angles between them, α, β, and γ. There are seven crystal systems for describing unit cells based on their shape. Each of the unit cells can be built up in one of 14 Bravais lattices to produce the crystal. Because the unit cell is chosen as the smallest unit that does not require rotation, the contents of the unit cell often display rotational symmetry. The rotational symmetry within the unit cell can restrict its shape so that only certain point group symmetry elements are associated with each of the 14 Bravais lattices (Figure 1.1). The smallest unit which possesses no crystallographic symmetry is known as the asymmetric unit. In addition to rotational symmetry, crystals can also possess screw axes, where the asymmetric unit is not only rotated around the symmetry axis, but also translated a fraction of the unit cell length. Screw axes are denoted as subscript numbers related to the fraction translation of the unit cell e.g. 2_1 is a two-fold rotation axis with a screw corresponding to half the unit cell length. The combination of Bravais lattices,

symmetry point groups and screw axes combine to give a possible 65 space group for non-centrosymmetric molecules such as proteins. The details of the symmetry and equivalent positions for these space groups, as well as other possible space groups for centrosymmetric objects, are found in Volume A of the *International Tables for Crystallography* (Henry and Lonsdale 1952).

Crystal system	Bravais lattices	Unit cell length (Å)	Angles (°)	Point groups
triclinic	*P*	$a \neq b \neq c$	$\alpha \neq \beta \neq \gamma$	1
monoclinic	*P C*	$a \neq b \neq c$	$\alpha = \gamma = 90, \beta > 90$	2
hexagonal	*P*	$a = b \neq c$	$\alpha = \beta = 90, \gamma = 120$	6, 622
trigonal	*P R*	$a = b = c$	$\alpha = \beta = \gamma \neq 90$	3, 322
cubic	*P I F*	$a = b = c$	$\alpha = \beta = \gamma = 90$	23, 432
tetragonal	*P I*	$a = b \neq c$	$\alpha = \beta = \gamma = 90$	4, 422
orthorhombic	*P I C F*	$a \neq b \neq c$	$\alpha = \beta = \gamma = 90$	222

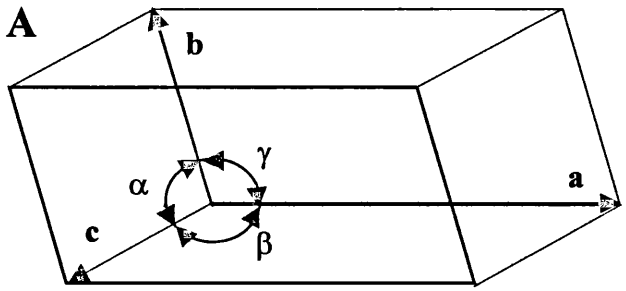


Figure 1.1: The table shows the unit cell dimensions that correspond to diagram **A** for the 7 crystals systems and the lattice types they can create. The types of lattices are *P*, primitive, *I*, body-centred, *F*, face-centred, *C*, base-centred and *R*, rhombohedral. The point groups denote the type and number of symmetry axes possible for these crystal systems.

When x-rays are shone at a crystal, the resulting diffraction pattern comprises many reflections, each arising from a plane of lattice points in the crystal. These planes include the faces of the unit cell but also other sets of parallel planes which are defined by where

they intersect the unit cell axes. The planes are given a notation (*hkl*), where each of these denotes into how many parts this set of planes splits each axis of the unit cell. Although there are many sets of planes though the crystal, for any one orientation of the crystal coherent diffraction can only occur from sets of planes whose spacing conforms to Bragg's Law. Bragg's law (Figure 2.1) shows that for any angle of incident radiation which hits the crystal, coherent diffraction only occurs from those sets of planes whose spacing allows a difference in path length between the waves to equal to an integral number of wavelengths (Blundell and Johnson 1976). This keeps the resulting diffracted rays in phase and therefore produces a coherent signal. Bragg's law is expressed mathematically as

$$2d \sin\theta = n\lambda ,$$

where d is the inter-planar distance, θ is the angle of the incident beam, n is the number of wavelengths and λ is the wavelength.

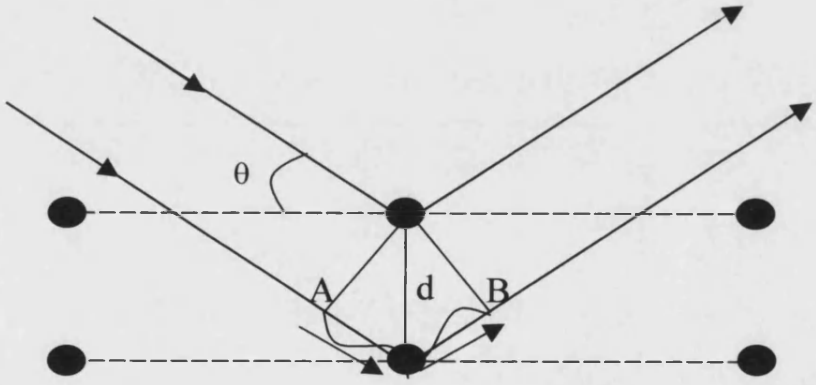


Figure 1.2: A schematic to represent Bragg's law. The extra distance travelled by the second wave, between A and B, must equal a whole number of wavelengths to produce coherent diffraction. θ is the angle of incident radiation and d the inter-planar spacing.

Because waves are only diffracted coherently from sets of planes in the crystal, the diffraction pattern observed for a single crystal is a coherent pattern of dots whose properties are related to the dimensions of the crystal. The diffraction pattern is the inverse of the crystal structure. The waves which are diffracted from planes in the real lattice of the

crystal, form dots, or reflections, in the reciprocal lattice of the diffraction pattern. The reciprocal nature of the diffraction pattern can be comprehended with reference to Bragg's law. Those sets of planes which are closer together will only give coherent diffraction at narrow angles of incidence, leading to more disperse spots on the diffraction pattern, whereas those planes which are further apart will give coherent diffraction at wider angles of incidence resulting in spots closer to the centre of the diffraction pattern. From this it follows that information about the overall structure of the unit cell contents will be more concentrated in the low resolution reflections at the centre of the diffraction pattern, as these originate at the wider spaced planes in the crystal. The details will be more concentrated in the high resolution reflections further out in the diffraction pattern, which originate at the more narrowly spaced planes.

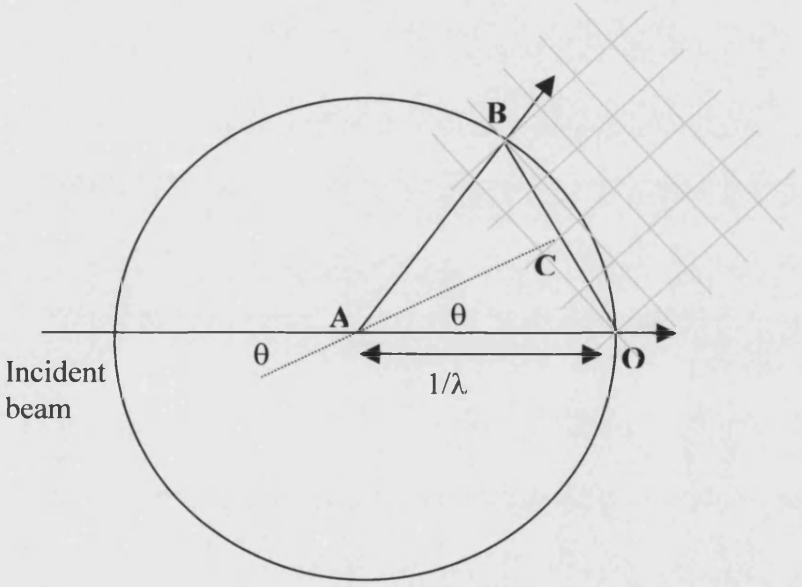


Figure 1.3 – The Ewald sphere construction for a set of planes that obey Bragg's law. A is the crystal centre, and O, which represents the point of exit for the non-diffracted beam, the origin of reciprocal space. The radius of the circle is $1/\lambda$. When the reciprocal lattice is rotated about O, the origin, point B in the reciprocal lattice intersects with the circle resulting in a reflection. This is because the magnitude of the vector OB, the distance between the origin of reciprocal space and the diffracted beam, equals $1/d_{hkl}$ where d is the lattice spacing. OC, which is $1/2d_{hkl}$, is equivalent to $1/\lambda \sin\theta$. Therefore, $1/2d_{hkl} = 1/\lambda \sin\theta$ which is equivalent to $\sin\theta\ 2d_{hkl} = \lambda$, is Bragg's law if n=1.

As discussed previously, for any one angle of incident radiation, coherent diffraction will only occur from some sets of planes in the crystal the conform to Bragg's law. This implies that diffraction patterns from many angles of incident radiation must be recorded to obtain all the information about the crystal. This can be easily achieved by rotating the crystal. This idea can also be described in reciprocal space by a sphere of reflection. The construction in Figure 1.3, shows that in reciprocal space there is a sphere of possible reflections waiting to happen when the crystal, and therefore the reciprocal lattice, rotates to bring them into view (Blundell and Johnson 1976). The radius of this sphere is the inverse of the real space wavelength. From this it follows that the shorter the wavelength of the incident radiation, the larger the sphere of reflection, allowing more of the potential reflections to be measured. Alternatively, there is more information in the diffraction pattern which is gained by access to the more closely spaced planes within the crystal.

The reflections that are recorded on the diffraction pattern actually represent a sum of waves diffracted from the atoms on the planes and are known as structure factors. All waves can be described as a sum of simple sine and cosine waves of the form

$$f(x) = F\cos 2\pi(hx + \alpha),$$

where $f(x)$ is the vertical height of the wave, F is the amplitude, h the frequency and α the phase. Alternatively, waves can be described as a combination of sine and cosine terms expressed as a complex number of the form

$$f(x) = F_h[\cos 2\pi hx + i\sin 2\pi(hx)],$$

or

$$f(x) = F_h e^{2\pi i\,(hx)}.$$

In this form, the phase is implicit in the combination of sine and cosine terms. The structure factor can therefore be described as a sum of three dimensional waves as

$$f(x,y,z) = \Sigma\, F_{hkl} e^{2\pi i\,(hx + kl + lz)}.$$

The contribution of individual atoms to the diffracted wave can also be written in this form and is known as the atomic structure factor. It is written as

$$f_{hkl} = f_j \, e^{2\pi i(hxj + kyj + lzj)},$$

where f_j is the scattering factor of the atom j, and depends on the size of the atom and x_j, y_j and z_j are the coordinates of the atom. The structure factor can therefore also be described as a sum of these atomic structure factors

$$F_{hkl} = \Sigma f_j \, e^{2\pi i(hxj + kyj + lzj)}.$$

The intensity of the measured reflection is equal to the squared structure factor amplitude, but contains information about the atoms contributing to the diffraction. The diffraction pattern in reciprocal space, which can be described as a sum of three dimensional waves, and the molecular structure in real space, which can also be described as a sum of three dimensional waves, are therefore related by a Fourier transform. Fourier transforms are functions which link functions of one variable with functions of another variable. The structure factors can therefore be described in terms of the volume of electron density at any point in the unit cell by the Fourier transform

$$F_{hkl} = \int_V \rho(x,y,z) \, e^{2\pi i(hx + ky + lz)},$$

where ρ represents the electron density at point x,y,z for a volume, V, of the unit cell. Alternatively, the function of electron density can be described as a Fourier transform of the structure factors as

$$\rho(x,y,z) = 1/V \sum_h \sum_k \sum_l F_{hkl} \, e^{-2\pi i(hx + ky + lz)}.$$

These two equations are therefore essential to retrieving information about the molecular structure from the measured intensities, but also for checking the correctness of the experimental model against the observed data.

1.3 Protein expression and purification

The technique of protein crystallography (Figure 1.4) is often regarded as starting at the point of expression and purification of protein for crystallisation. Good data requires good crystals and good crystals require a good supply of highly pure (>95%), soluble, homogeneous protein. The use of recombinant expression systems have been a valuable tool in the production of large amounts of proteins which are relatively scarce in nature. This has lead to standard procedures where proteins from many different organisms are cloned into vectors for expression in *Escherichia coli* (*E. coli*). These vectors are designed

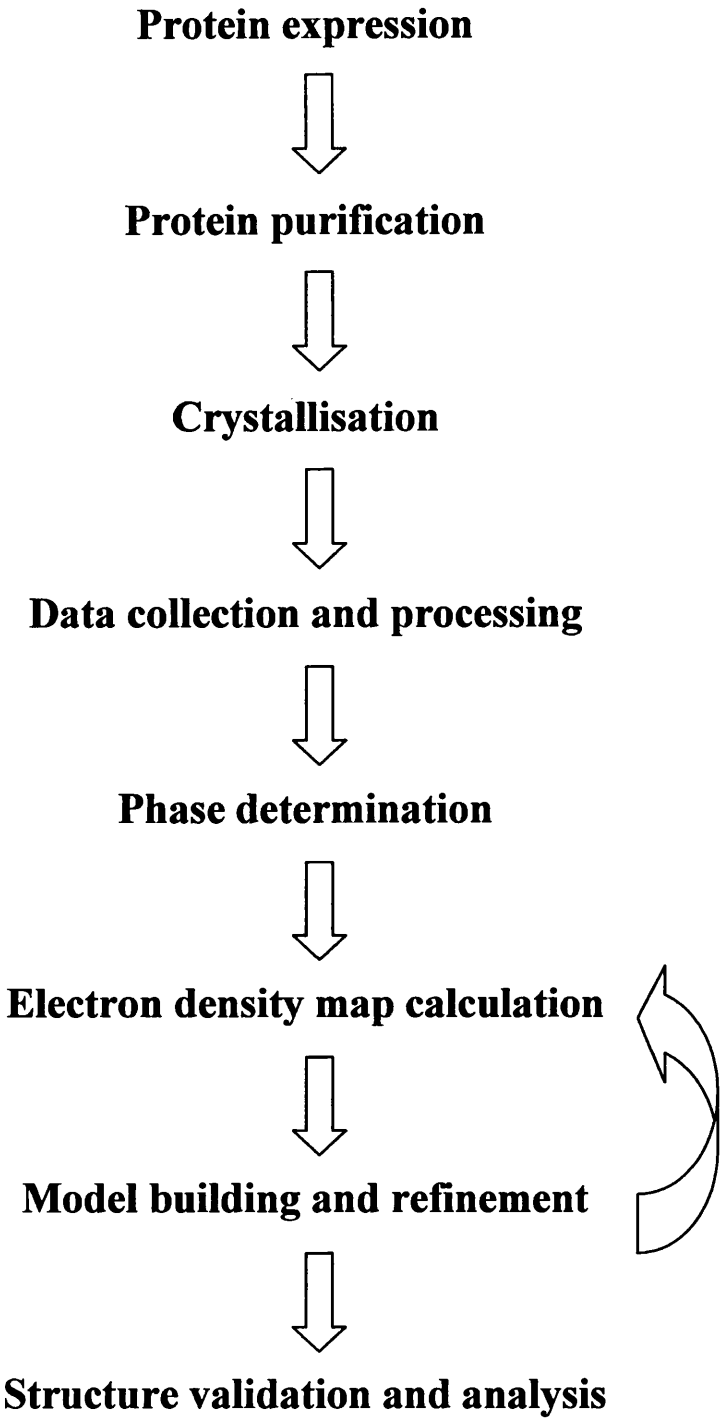


Figure 1.4: A flow chart showing the steps in structure determination by protein crystallography.

with high expression promoters, which are often inducible. They allow levels of protein expression much greater than the native level that are clearly visible by SDS-PAGE. Unfortunately, bacterial recombinant expression systems often result in improperly processed proteins, in the case of animal and plant proteins, or insoluble protein. Incorrect processing of proteins can be circumvented by alternative expression systems such as yeast cultures, insect cells, mammalian cells, or even plants, although these systems often cannot produce the same quantities of protein as bacteria. For some cases, *E. coli* can be engineered to express the tRNAs it lacks to increase expression and solubility of the recombinant protein. The use of *E. coli* proteins as expression tags such as glutathione S-transferase (GST) (Smith 2000) and maltose binding protein (MBP) (Sachdev and Chirgwin 2000) can also lead to greater solubility of the protein. These tags also aid purification by the use of glutathione and amylose affinity columns respectively. Another affinity tag used for purification is the His tag (Bornhorst and Falke 2000), comprising a short chain of histidine residues, which is purified with divalent metal ion columns. This tag does not help the solubility of the protein but is useful in studies where cleavage of the tag is inconvenient. Alternatively, insoluble proteins can also be refolded from inclusion bodies using a denaturant followed by dialysis or drop-wise addition into a suitable buffer. If optimised, this protocol often acts as a good purification step, removing the soluble protein.

1.4 Crystallisation

Once suitable protein has been produced, it is then persuaded to crystallise by mixing with various precipitants. There are two commonly used methods for crystallisation. The first is the direct mixing of protein and precipitant under oil. This does not allow for evaporation of the precipitant and gives a quick assessment of the proteins solubility in the various precipitants tried. This technique is becoming more common with large scale crystallisation trials where robots can set up many very small drops with ease, conserving protein. The more traditional method, still widely used by those who set up their crystallisations by hand, is the vapour diffusion technique. In this technique, a small volume of protein and precipitant are mixed together and suspended over a reservoir of the precipitant in an air tight system. Water vapour can then diffuse from the drop into the reservoir, concentrating the protein and allowing crystallisation. Despite much research into crystallisation and characterisation of the crystallisation process, most crystallisation

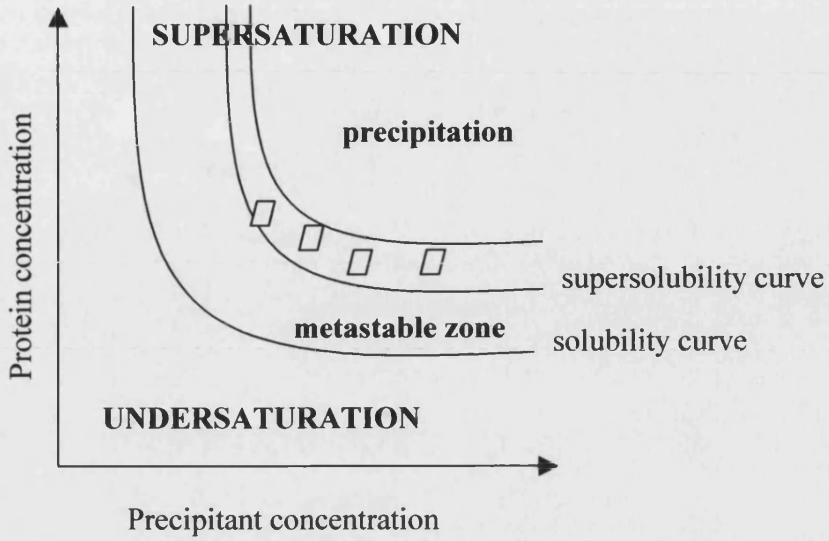


Figure 1.5: Diagram showing the change in protein saturation for increasing precipitant and protein concentrations. Below the solubility curve the protein is undersaturated and cannot crystallise. The metastable zone is where the protein saturation is not high enough to nucleate crystals but will allow growth of seeds. Above the supersolubility curve, there is too much protein to stay in solution and the excess protein will either nucleate crystals or precipitate. The ideal protein concentration is just high enough to nucleate crystals which will then grow in the metastable zone due to the decrease in soluble protein concentration during crystal growth.

attempts are characterised by a trial and error method. In some cases previous knowledge of the solubility of the protein in different salts or buffers may be used to design a crystallisation trial, but more often a commercial screen of widely used precipitants is used as a starting point. If no preliminary crystals are produced, however, the protein solubility in various precipitants must be used to design further screens.

The solubility of a protein depends on several factors of which the main ones are its concentration, the pH of the solution, the concentration of any precipitants and the temperature (McPherson 1985). For crystals to grow, these factors need to conspire to

supersaturate the protein in the solution without precipitating it (McPherson 1990). The effect of protein and precipitant concentration on the solubility can be demonstrated schematically by a simple graph as shown in Figure 1.5. Therefore, the aim of the crystallisation trial is to optimise all these factors to give good quality crystals.

Although there is some guess-work in determining the type and concentration of precipitant to use, it is well known that there are two stages to crystal growth, nucleation and growth. However, the conditions required for nucleating crystals are not always the same as those which encourage growth of a large crystal from the nucleus. This has lead to the development of a technique known as seeding, where small crystals nucleated in one condition, are transferred to a slightly different condition to be allowed to grow further. This technique can also be used when a condition that nucleates well, cannot grow the crystals due to protein depletion following too much nucleation. There are three main ways of seeding: macroseeding, microseeding and streak seeding. Macroseeding involves the transfer of one small crystal, or 'seed', into a new drop, sometimes after a 'wash step' to get rid of any impurities that may be preventing growth on the surface (Thaller *et al.* 1985). Microseeding avoids that problem by transferring small fragments of crystals which have been crushed and diluted with the crystallisation solution. The dilution step allows control of how many 'seeds' are added to the drop (Fitzgerald and Madsen 1986). Streak seeding involves transferring seeds into a new drop by touching an existing crystal, or crystalline precipitate, with a cat's whisker, or similar, and then running it through the new drop (Stura and Wilson 1991). It can be harder to control the number of seeds in this method but works well for some crystals.

1.5 Crystal manipulation

The mounting of a crystal for data collection depends on whether the crystal is going to be frozen. For data collection at room temperature, the crystal must be sucked from the crystallisation drop into a thin glass capillary, with some of the crystallisation solution, or 'mother liquor'. A drop of mother liquor in the capillary with the crystal is essential to stop the crystal drying out, due to the high water content of protein crystals (Blundell and Johnson 1976). This capillary is then sealed at both ends with wax, ready for mounting on the goniometer head. It is becoming more common, however, to freeze crystals at 100K for diffraction as this can reduce the amount of radiation damage the

crystal receives, allowing more data to be collected from a single crystal. This is particularly useful for anomalous dispersion techniques if several wavelength data sets can be collected from one crystal to avoid non-isomorphism. However, not all crystals can be successfully frozen without disruption to the crystal lattice. Crystals that are frozen often need to be soaked in a cryoprotectant first. This reduces the amount of water picked up with the crystal and also prevents icing of the solvent in the crystal on freezing. Both excess water and ice diffract x-rays to produce noise in the diffraction pattern. Cryoprotectants are usually viscous liquids such as glycerol, oils, or high concentrations of polyethylene glycol (PEG). The crystal, held by a nylon loop, is soaked briefly in the cryoprotectant before being mounted on the goniometer head where it is flash-frozen in the cold nitrogen stream used during data collection.

1.6 X-ray sources

The X-rays used for protein crystallography usually come from one of two sources. Crystallography laboratories often have rotating copper anode tube as a 'home' source. Electrons from a cathode are accelerated towards a rotating, water-cooled, copper anode, resulting in emission of Cu-Kα radiation at a fixed wavelength of 1.54Å. Home sources are ideal for characterising crystals and collecting data for crystals that diffract well. Synchrotron sources, however, have many advantages of shorter wavelength radiation, tuneable wavelength radiation, greater brilliance and greater flux (Helliwell 1997). Synchrotron radiation is produced when electrons are accelerated near the speed of light by a magnetic field. The forced curving of the electron path as they are accelerated around the large storage ring, causes a loss of energy resulting in the emission of electromagnetic waves. This electromagnetic radiation, emitted at a tangent to the ring, is then guided and focussed for use in experiments. Because of the complete range of electromagnetic radiation produced, the experimental stations can be designed to offer a range of X-ray wavelengths for diffraction experiments. This not only allows very high resolution experiments, but also the ability to choose wavelengths close to the absorption edge of heavy metals for anomalous scattering experiments.

1.7 Data collection

There are many factors to be considered in data collection to achieve the highest possible resolution data set for the crystal. The diffraction limit depends on the order of the

crystal and the wavelength of the x-radiation being used. The increasing use of synchrotron radiation at shorter wavelengths under 1Å has lead to more high resolution structures. However, the intensity of synchrotron radiation which allows a better signal to noise ratio, also brings problems of radiation damage. Radiation damage can be reduced by freezing of the crystal, as mentioned above, but has also led to an improvement in detector systems. The quicker the data can be recorded for each orientation of the crystal, the less time the crystal is exposed in the beam. A commonly used detector at synchrotrons is the charge-coupled device, or CCD, detector. CCDs accumulate charge on the pixels of the detector where the radiation hits. These charges are systematically passed via fibres to an amplifier before reading. The CCD detector has the advantage of accurate recording of the position and magnitude of the reflections, which are continuously read, making data collection fast and accurate.

The advances in cryo data collection and fast detectors combined with automation of the data processing and scaling software allow highly redundant, mostly complete data sets to be collected with ease for most good crystals. HKL2000 (Otwinowski and Minor 1997) can process the data simultaneously with data collection. This provides immediate assessment of the quality of the data as well as the completeness of the data set, allowing alterations to be made to the data collection strategy if necessary. Rapid deterioration of a crystal can be spotted allowing the crystal to be changed. Also, completeness of data sets for very mosaic crystals can be monitored. Mosaicity, or disorder, in the crystal can cause each Bragg reflection to be recorded in more than one position, making the data harder to process. In the case of high mosaicity, high redundancy in the data is required to be able to scale a complete data set.

Although the data collection can be monitored, several important factors should be considered before starting a data collection run. Initial images of the crystal should be used to optimise the exposure time, depending on the strength of diffraction, the crystal to detector distance, depending on the diffraction limit of the crystal, the probable number of images to be collected, depending on the symmetry of the space group, and the angle through which the crystal is rotated, depending on the size of the unit cell (Dauter 1997). For example, most reflections for the space group P222 can be collected with a 90° rotation of the crystal, although more images should be collected to get good redundancy to reduce errors.

1.8 Data processing and scaling

The choice the unit cell within the crystal lattice can be arbitrary, but is usually chosen as the smallest volume unit cell that displays obvious rotational symmetry and retains perpendicular lattice translations. The crystal lattice spacings are calculated as the inverse of the reciprocal lattice of the diffraction pattern by software which then statistically ranks possible lattices and unit cells. It is practice to choose the best ranked highest symmetry lattice to process the data in. The space group is chosen during scaling based on systematic absences. The procedure of processing the data involves the indexing all of the reflections from planes in the chosen unit cell. Following indexing, the data is scaled so that the backgrounds from different diffraction images match (Leslie 2001). Scaling can also combine data sets from more than one crystal, as long as the crystals are isomorphous, i.e., have the same unit cell dimensions.

The output from the programs that scale the data display important statistical information about the quality of the data. The first of these is the data completeness. For a given space group, at a particular resolution, the number of possible reflections from planes in the crystal can be calculated. This is then compared to the number of unique reflections recorded in the data collection and the completeness calculated as a percentage. A complete data set, which is usually easily achievable with modern x-ray sources and fast detection systems, is ideal, and can make the structure solution more straight forward. It is particularly necessary for anomalous scattering phasing methods to ensure the anomalous signal is as clear as possible. For less than ideal situations, e.g. where the crystals are very mosaic, a lower level of completeness can still allow a more careful structure solution.

Another measure of the quality of the data is R_{sym}. R_{sym} measures the ratio of the mean difference between symmetry related intensities that should be the same and the mean magnitude of the intensities. This value depends on the order of the crystal but may also reflect inaccurate detecting of the reflections. As many of the reflections are measured more than once in data collection, another value often quoted is R_{merge}, which averages the intensities and differences of the other equivalent reflections as well. A low percentage value of R_{merge} indicates accurately recorded data, but it is important not to ignore less accurate reflections at the expense of the completeness. A value of 5% for R_{merge} is very

good. For higher resolution structures, or more redundant data, values of up to 15% may be acceptable, as R_{merge} increases with redundancy.

The resolution to which a data set can be considered reliable, depends on the completeness, R_{merge}, and also the signal to noise ratio, known as $I/\sigma(I)$ (Dauter 1997). This is the ratio of the mean reflection intensity to the standard deviation of the intensities. All of these statistics are calculated for resolution bins as well as the entire data set and in the case of $I/\sigma(I)$, the figure for the highest resolution bin will show whether spots for that resolution can actually be seen. The Wilson B factor for the data is also linked with resolution. The Wilson B factor, or temperature factor, gives an indication of the disorder of the crystal by the comparison of the decrease in the intensity of reflections with increasing resolution. To take into account the decrease in the contribution of any particular atom at higher resolution, the mean intensity for a given resolution is divided by the mean intensity for the average protein. A graph of the log (I/f^2) against $(\sin\theta/\lambda)^2$, where I is the mean intensity, f is the mean protein and λ is the wavelength, gives a straight line with a gradient equalling –B in $Å^2$. It is practise to set the initial atomic B factors for your model to this value.

When data are being assessed, all these statistics should be considered as the relationship between them often allows one to be improved at the expense of the others. For example, the completeness of the data can usually be improved by limiting the resolution, with loss of the higher resolution reflections. The scaling of the data therefore needs to combine these factors sensibly to produce the highest resolution, best quality data set.

1.9 Phasing methods

The necessity of collecting unfocussed x-rays as a diffraction pattern causes the loss of phase information and has been termed 'the phase problem'. In order to calculate the molecular transform from the diffraction pattern the phases need to be estimated. Currently, there are three techniques in use and crystallographers choose one or a combination depending on the appropriateness for the problem. Two of these methods provide 'novel' phases, for molecules for which no structural information is known whereas the third requires the use of a homologous model.

1.9.1 Novel phasing methods

Multiple Isomorphous Replacement- Multiple Isomorphous replacement (MIR) is the classical method for solving new structures which utilises phase information from the comparison of structure factors from native and derivative crystals, soaked in heavy metals (Green *et al.* 1954). Because atoms contribute most strongly to structure factors from planes which intersect their positions, the uniform addition of heavy atoms, with larger scattering factors, at certain points in the unit cell gives rise to an increase in the amplitude of those structure factors. If the native and the derivative crystals are isomorphous the resulting intensities can be compared with the native intensities to retrieve phase information for these atoms. This phase information can then be used to position these heavy atoms in the unit cell to provide a rough framework for the model to be built into. To determine the phases as accurately as possible, several derivatives are often necessary, requiring the careful production of many isomorphous derivative crystals.

Anomalous Scattering- The other technique for novel phase estimation uses the absorption of x-radiation by heavy atoms to derive information about phases (Hendrickson *et al.* 1985). All diffraction patterns are centrosymmetric as diffraction from each plane gives 2 reflections, *hkl* and *–h-k-l* which are equal in intensity. This is known as Friedel's law. All atoms absorb radiation at specific wavelengths which correspond to the required energy for electron excitation. For certain heavy atoms, these wavelengths correspond to X-radiation wavelengths suitable for crystallography. If the x-rays used for the experiment are tuned to this absorption edge, the absorption of the radiation results in an observable breakdown in Friedel's law known as anomalous scattering, which can be used to retrieve phase information for these atoms. This technique can combat the problem of non-isomorphism of heavy atom soaks as the same frozen crystal may last long enough to record native (at a low-absorbing wavelength) and anomalous data sets. Another advantage of this method is the option of using seleno-methionine, with an absorption edge at 0.98Å, which can be incorporated during expression of the protein. Bacterial cultures grown in seleno-methionine media can be used to produce recombinant proteins where methionine is uniformly replaced with seleno-methionine, which circumvents the need for crystal soaking. However, anomalous scatterers can be added by crystal soaking which requires optimisation by trial and error, as for MIR, to produce good diffracting, isomorphous,

derivative crystals. This technique is also particularly useful for metal containing proteins such as metalloproteases.

The use of a combination of both these methods to estimate phases is also increasingly common. Phase information from the increased scattering factor and the anomalous scattering of heavy atoms can be combined in a technique referred to as Single Isomorphous Replacement with Anomalous Scattering, or SIRAS. In practise, there are now a good range of phasing and density modification programs that with the use of excellent data collection technique and careful refinement, structures can be solved with very minimal phase information.

1.9.2 Molecular replacement

If the structure to be solved has over 30% sequence identity with another protein, this usually indicates that they share a common fold. Molecular replacement takes advantage of the reversible nature of Fourier transforms to calculate phases from a known model to use as an estimate for the structure to be solved. The proposed model is unlikely to share the same space group or orientation as the unknown structure so the molecular replacement program is required to find a rotation matrix and translation vectors which relate the known model to the unknown structure. This is achieved by moving the known model through a random unit cell with the unknown structure in. In practice, however, to find the six variables is a computationally demanding task and so the search is split into finding the orientation first and then the translation.

This is commonly done using Patterson functions. Whereas the Fourier transform relates the intensities and phases of the reflections to the model structure, the Patterson function relates the squared intensities of the data to a set of vectors between the atoms. Therefore, a Patterson function can be accurately calculated for both the known model and the unknown structure.

The Patterson function can be written as

$$P(\mathrm{u}) = \Sigma \mid F_h \mid^2 \exp^{-2\pi i h u},$$

which is the same form as the electron density equation but using squared amplitudes rather than a phase term in the exponential. As the Patterson function describes the relation of the atoms within the structure to themselves, if the known model and the unknown

structure are indeed similar in shape, they should produce similar Patterson maps which can then be rotated until they match (Rossmann and Blow 1962). For whole proteins, however, the Patterson map is very complicated and the parameters of the search often need careful planning.

In order to find a good match where the two structures only share low sequence identity it is often necessary to perform the rotation search at low resolution This prevents the details of the two structures, which are likely to be different, confusing the matching of the overall fold. The model used is also an important factor and it is often helpful to delete flexible termini or different sized loops which enhance the differences between the two structures. Additionally, the sphere radius over which the functions are compared can affect the success of a match. The Patterson map from the unknown crystal structure will also contain vectors which relate symmetry related molecules in the unit cell, known as cross-vectors. The sphere radius therefore needs to be large enough to include enough intra-molecular vectors so a good match can be found, but not too large, that the match is overwhelmed by the noise of the intermolecular vectors. The translation function also uses Patterson functions and works in a similar way. In this case, the intermolecular vectors of the unknown structures are overlapped with generated intermolecular vectors, of the same symmetry, for the model.

The hardest part of finding a molecular replacement solution is actually working out at which coordinates the known model superimposes on the unknown structure. The problem of identifying a correct solution for molecular replacement occurs due to the way the matches are ranked statistically. The program AMORE (Navaza 1994) ranks possible rotations by correlation coefficient and R factor. The correlation coefficient compares the position of the peaks in the Patterson functions and their signs within a volume of the model and scales the results to -1 to 1. The R factor is the ratio of the difference between the observed and calculated amplitudes to the size of the observed amplitudes and can be expressed mathematically as

$$R = \frac{\Sigma \, | \, |F_{obs}| - |F_{calc}| \, |}{\Sigma \, |F_{obs}|} \quad .$$

However, the solution with the best statistical scores is not always the correct one, and solutions need to be checked to be able to generate proper, non-clashing, symmetry related molecules on the graphics computer. The program Molrep (Vagin and Teplyakov 1997) checks the symmetry clashes as well so is therefore more likely to offer the correct solution.

If there is more than one molecule in the asymmetric unit, molecular replacement will have to be repeated for each of the other molecules. The knowledge of the number of molecules in the unit cell comes from the Matthews coefficient (Matthews 1968) which, from the size of the protein and the unit cell, estimates the solvent content of the crystal for different numbers of molecules in the unit cell. These solvent contents are then compared to the average solvent content of protein crystals to determine how many molecules per asymmetric unit there are.

1.10 Electron Density Maps, density modification and model building

Once the rotation and translation operations, or the heavy atom positions, have been found the model phases can be combined with the observed amplitudes from the diffraction pattern. These structure factors can then be refined against the model structure. If the MR solution or the heavy atom positions are correct, the R-factor, which is the ratio of the difference between the computed amplitudes for the model and the observed amplitudes, to the observed amplitudes, should go down. These structure factors can then be used to generate an initial map, which if the phase information is good enough, should be interpretable. There then follows many round of model building and refinement to generate a model which represents the observed data as closely as possible.

One of the skills required by crystallographers is map interpretation (Jones and Kjeldgaard 1997). The electron density equation generates values for electron density for each point, $\rho(x,y,z,)$, in the unit cell, from the structure factors. These figures are then displayed as three dimensional contour maps on a graphics computer using a program such as O (Jones *et al.* 1991). In order not to bias the map to the model structure-factor amplitudes, these are normally subtracted from a multiple of the observed amplitudes e.g. a $2F_o$-F_c map is the sum of twice the observed structure factor amplitudes, minus the

calculated amplitudes, multiplied by the phase term. This map is used to represent the positive, continuous electron density of the model. F_o-F_c maps are also generated and these provide both positive peaks, where there is something in the data which is lacking in the model, and negative peaks, where there is something in the model which is not in the data. This map, along with the amino acid sequence of the protein, is essential in improving the model so that the calculated model amplitudes more closely resemble the observed amplitudes. Both the $2F_o$-F_c map and the F_o-F_c maps are contoured to a set level for viewing in the graphics, but local adjustment of these levels is often useful to see where the peaks are strongest to help interpret the map. Also, for parts of the model which are uncertain, the use of omit maps can be very useful. These maps omit all the structure factors for the uncertain bit of the model out of the map calculation to produce a map without model bias for that area.

For structures solved by molecular replacement, the initial model is usually reasonably good and gives a map which is interpretable in most places. Initial maps for structures solved by multiple isomorphous replacement or anomalous scattering, however, are not always as easy to interpret. For these maps in particular, the technique of density modification is very useful. Density modification takes general knowledge about protein structures and electron density maps not found in the structure factors to improve the maps (Zhang *et al.* 1997). The first of these is solvent flattening. Protein crystals often contain 50% or more solvent which is disordered in the crystal and therefore does not produce strong peaks in the map, unlike the protein. Solvent flattening attempts to define the boundary between the protein and the solvent mass and then assign the solvent to a fixed level. This electron density distribution can then be used to calculate new structure factors, the phases of which can be combined with the previous phases to produce a better estimate, and therefore a better map. This effectively makes the protein part of the map stand out more from the bulk solvent. Histogram matching is another technique that improves the protein part of the electron density to improve the phases. The electron density can be represented as a histogram with peaks where protein regions lie. From knowledge of previously refined structures, it is known that these histograms have a usual shape for protein structures at particular resolutions. In this technique, the histograms are fitted to this expected shape, improving the electron density and therefore the phases. Another essential density modification technique is the use of non-crystallographic symmetry. Where there is more than one molecule in the asymmetric unit, providing the symmetry

axes which relate these molecules are known, the electron density peaks for these equivalent regions can be averaged to cancel out noise from the solvent and poor phases. This technique has been particularly powerful for structures with multiple copies (e.g. virus structure solution).

Model building itself is often carried out with the program O (Jones *et al.* 1991) on the graphics computer. This program allows the electron density to be displayed as a contour map and a stick representation of the model coordinates to be manipulated to fit the map better. The new coordinates from the manipulated stick representation can then be output for refinement.

1.11 Refinement

Between rounds of model building in real space, reciprocal space refinement is also used to improve the model. Refinement is the optimisation of a function, to find the global minimum, by changing the parameters of the model. The function to be optimised in crystallography is the agreement between the observed structure factor amplitudes and the calculated model amplitudes. The model parameters include the position, occupancy and B factors of the atoms. The function used for this refinement, as well as the method of optimisation, varies between refinement programs. The traditional function used as implemented by SHELX (Sheldrick and Gould 1995; Sheldrick and Schneider 1997) is the Least Squares (LS) function which is the sum of the weighted differences between two sets of observations (the observed and calculated amplitudes). The limitation of least squares is that it assumes the errors in the model are independent and normally distributed and so mainly works best on models where the phase information is already good. An alternative to Least Squares is the Maximum Likelihood (ML) function implemented in Refmac (Murshudov *et al.* 1997) and CNS (Brünger *et al.* 1998). The ML function finds the set of parameters which are the most probable to produce the experimental results seen. The advantage of ML is that, although it still assumes that all errors are independent, it models the errors better and does not assume they are normally distributed. The third option is energy minimisation, also implemented in CNS, for which the target is the model with the lowest calculated energy.

As well as differing in the target function, programs also differ in their method of optimisation, which often involves derivitising the function. An example of optimisation using no function derivative is simulated annealing, implemented by CNS (Brünger and Rice 1997). This technique optimises to the lowest energy model. It has a large radius of convergence and the advantage of overcoming local minima by simulated slow cooling of the model from a high temperature. Derivative function methods include conjugate gradient method, also used in CNS and full matrix minimisation implemented by SHELX. Derivative optimisation allows a faster rate of convergence of the function, but often at the expense of a smaller radiance of convergence, making entrapment in local minima more likely.

To avoid model bias and entrapment in local minima, it is often helpful to start refinement at a lower resolution and increase this resolution in one or more steps. This ensures the gross features of the model are correct first and that wrongly assigned details do not bias the model. Using restraints and constraints in refinement is also important, especially at low resolution. Model parameters can only be optimised if there is enough data that the refinement is over determined. Adding restraints increases the observation to parameter ratio. Therefore, a good technique is to restrain the geometry of the protein tightly until the phases are more accurate as distortions in local geometry cannot be assigned without good phases. Restraining non-crystallographic symmetry partners together is also particularly useful and can average out noise in the phases. Initial refinement is usually rigid body refinement where the protein is refined as a whole followed by optimisation of atom positions. If the data is of a high enough resolution, water molecule positions and protein B factors can also be refined.

The refinement process is usually monitored by two parameters, R_{cryst} and R_{free}. The R_{cryst} (also just known as R) is calculated as the sum of the differences between the modulus of the observed and calculated structure-factor amplitudes divided by the sum of the observed structure-factor amplitudes as discussed previously. This is a measure of how well the model fits the data. The R_{free} is calculated as the R_{cryst} but for a randomly selected percentage of reflections known as the test set which are not used in the refinement (Brünger 1992). This acts as a cross validation tool to check that the refinement is not biasing the model away from the data. As cycles of real space modelling and reciprocal space refinement proceed, both R factors should decrease as the model structure converges

with the observed data. Towards the end of refinement structured solvent molecules can be modelled and these reduce the R factors further. The quality of the model should not just be judged by the R factors though and it is important that the structural geometry is good, and that the map appears to be interpreted as far as possible.

1.12 Structural validation

Although R factors give an indication of the quality of the model, it is important to use additional measures to validate the model (Kleywegt and Jones 1997). The most common of these is the Ramachandran plot (Ramachandran *et al.* 1963) which displays the peptide phi and psi angles for all residues in the model. It is important that over 90%, and preferably, 100% of the residues are in the 'allowable' regions. This plot gives a good estimate of the geometric accuracy of the model. Root mean square deviations from ideality can also be calculated for bond angles and lengths. It is important that models adhere to these geometric restraints unless they are of a high enough resolution (>1Å) that there is sufficient data to refine them. Another important factor when analysing structures is the B factor as although a low B factor does not always mean a good model, or a well placed side chain, a high B factor can indicate a problem with the model or high mobility which cannot be modelled. Often density is missing for regions with high B factors. It is important when analysing structures to consider all these indicators as none of them are sufficient in isolation. The program PROCHECK (Laskowski *et al.* 1993) is useful for these types of validation. Once a correct model is achieved, it can then be analysed for its biological relevance and the coordinates deposited in the Protein Data Bank (PDB).

Chapter 2

ADP-ribosylating toxins and Botulinum neurotoxins

2.1 Introduction

The results in this thesis cover the C3-like exoenzymes (ADP-ribosylating toxins), their interactions with Rho GTPases and also the botulinum neurotoxins. Therefore, this introduction starts with a brief structural overview of the other ADP-ribosylating toxins as well as the Rho GTPases. Crystal structures of Rho GTPase complexes are also described as a background to how they interact with other proteins. These brief overviews are followed by a more in depth literature review of the best characterised C3-like exoenzyme, C3 from *Clostridium botulinum*. Finally, a brief introduction to botulinum neurotoxin structure and function is given. Brief introductions to later chapters will build on this background information to present the context of the work described as relates to the current literature.

2.2 ADP-ribosylating toxins

ADP-ribosylation is a common mechanism for many bacterial toxins. These toxins catalyse the cleavage of NAD with transfer of ADP-ribose to the target protein and release of nicotinamide (Figure 2.1). Bacterial ADP-ribosylating toxins (ADPRTs) are grouped into several structural and functional classes. All structures of ADPRTs solved so far reveal a similar enzymatic domain but are classified by their cell binding domains (for reviews

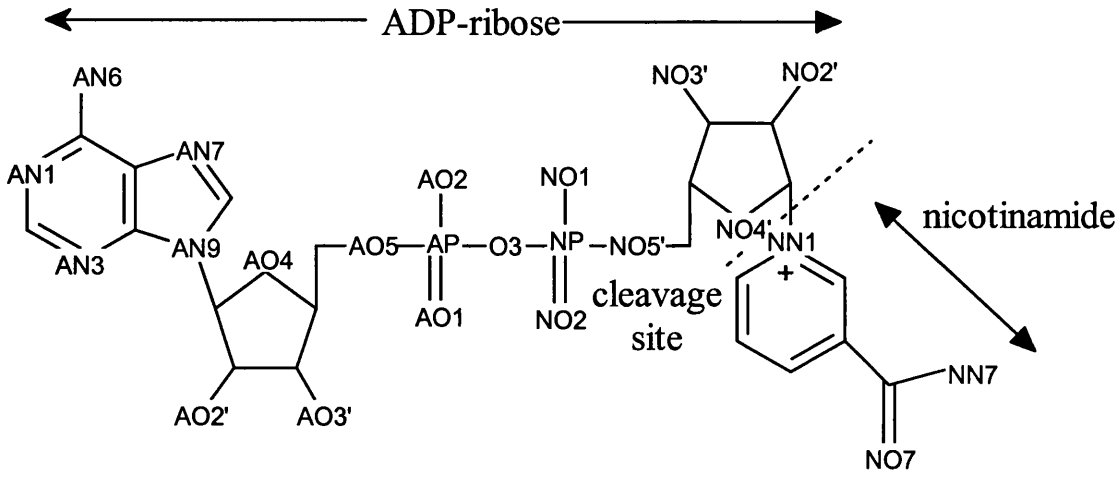


Figure 2.1: Schematic picture of NAD showing the cleavage site during ADP-ribosyltransfer.

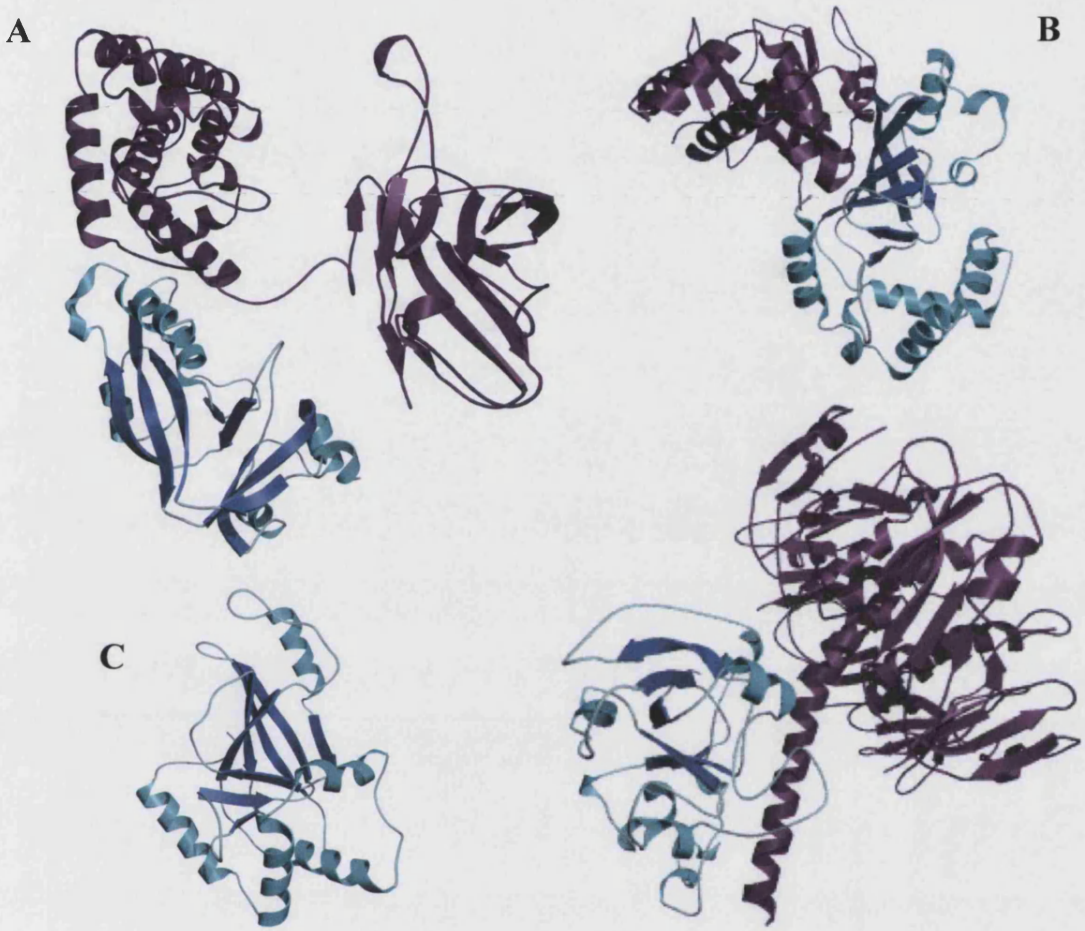


Figure 2.2: The cell binding (pink) and catalytic (blue/green) domains of the ADP-ribosylating toxins. **A**, Diphtheria toxin (PDB code 1SGK) (Bell and Eisenberg 1997), **B**, Catalytic component of the Iota toxin (PDB=1GIQ) (Tsuge *et al*. 2003), **C**, C3 exoenzyme (PDB=1G24) (Han *et al*. 2001) and **D**, Cholera toxin (PDB=1XTC) (Zhang *et al*. 1995).

see (Passador and Iglewski 1994; Krueger and Barbieri 1995) (Figure 2.2). The first structural group, which is characterised by Diphtheria toxin (DT) and *Pseudomonas aeruginosa* Exotoxin A (ETA), are synthesised and released as a single polypeptide but are later cleaved to form two domains linked by a disulphide bridge. These two domains comprise an A, or ADP-ribosylating domain and a B, or cell binding domain (Figure 2.2). Both toxins ADP-ribosylate elongation factor 2 (EF-2), inhibiting protein synthesis. Cholera toxin (CT) and *Escherichia coli* heat-labile enterotoxins (LT) characterise a second group of ADPRTs which ADP-ribosylate the small heterotrimeric G-proteins. These toxins also have A and B subunits but with the stoichiometry $A:B_5$. The binary toxins, which ADP-ribosylate actin, are characterised by *Clostridium botulinum* C2 toxin, *Clostridium perfringens* Iota toxin and Vegetative Insecticidal Protein (VIP) from *Bacillus cereus.* These toxins are synthesised as two separate ADP-ribosyltransferase and cell binding proteins which are not linked. Finally, the C3-like exoenzymes, characterised by C3 from *Clostridium botulinum*, ADP-ribosylate Rho GTPases and comprise a single enzymatic subunit with no cell binding domain. The groups of toxins can also be split into two families based on their catalytic residues. All ADP-ribosyltransferases utilise a conserved glutamate for catalysis. However, the CT-like, binary and C3-like toxins also require a glutamate/glutamine adjacent but one of the N-terminal side (Han and Tainer 2002) (Figure 2.3). This residue is associated with substrate recognition as it is a glutamine amongst the asparagine-ribosylating C3 exoenzymes (Figure 2.3) but a glutamate amongst the arginine-ribosylating binary and CT-like toxins. The DT-like toxins, like the eukaryotic poly ADP-ribosyltransferases (PARPs) lack this residue.

Figure 2.3: Sequence alignment of the C3-like and binary toxins. Residue numbers are from the first residue of the mature protein except for C3bot1 and 2 where they are labelled as the crystal structure. Residues conserved in all toxins are shown in red, those just conserved in C3-like exoenzymes in pink and those in the four binary toxins shown, in blue. The YXSTX motif is highlighted blue, the phosphate-nicotinamide (PN) loop, in grey and the ADP-ribosylating toxin turn turn (ARTT) loop in yellow. The arginines conserved in C3-like toxins are labelled above, 1-4. 2 and 3 are conserved in the binary toxins. Y246 from Iota toxin is labelled with a *. Alignment done in T-Coffee (Notredame *et al.* 2000).

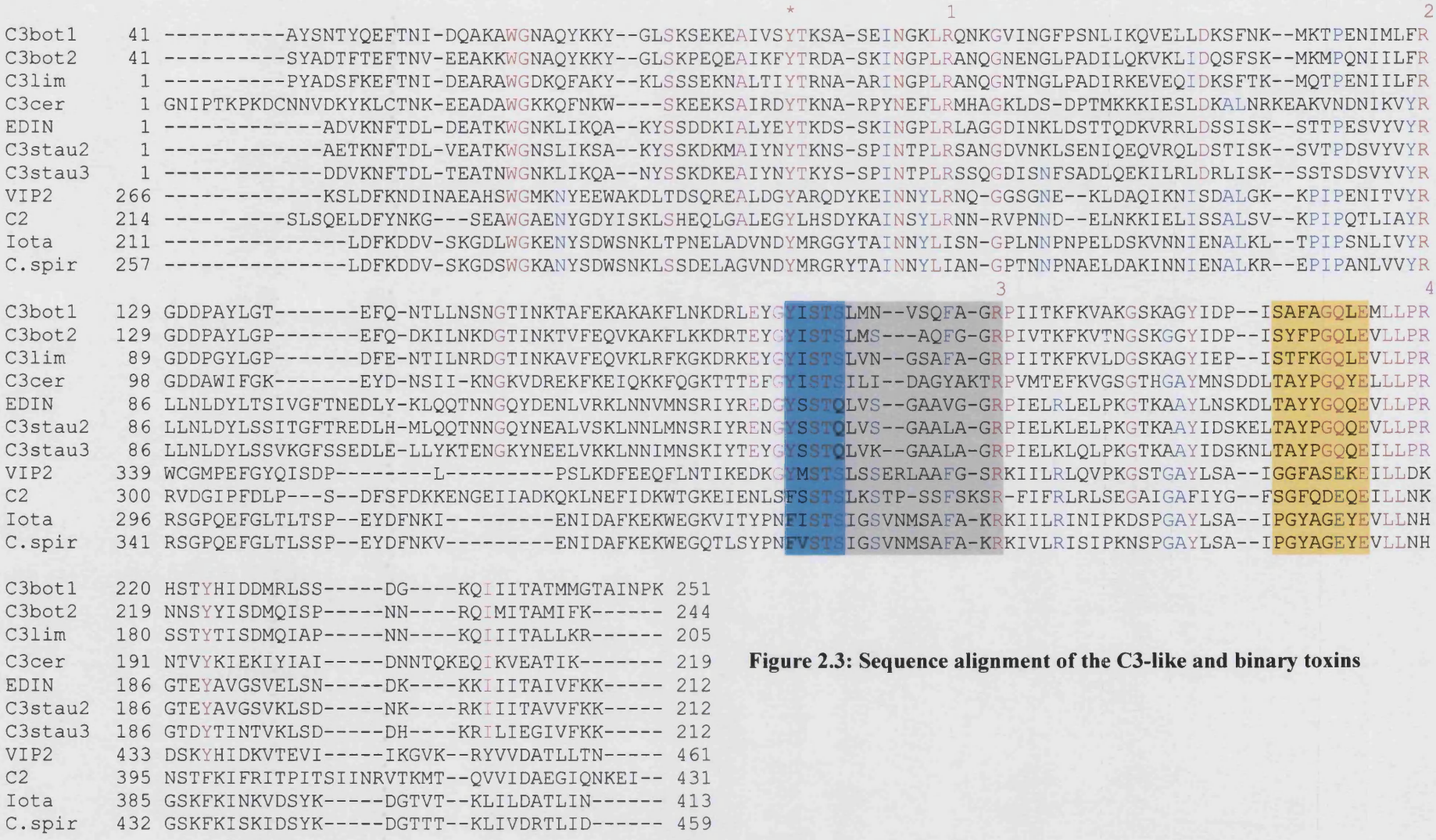


Figure 2.3: Sequence alignment of the C3-like and binary toxins

The first ADPRT structures solved, ETA and DT, revealed the ADP-ribosylating domain comprised two perpendicular β-sheets surrounded by α-helices (Figure 2.4). The active site resides in a cleft at the junction of the two β-sheets between one of the β-sheets and helix 3 (Choe *et al.* 1992). It is not similar to the Rossman fold binding cleft of dehydrogenases where NAD/NADH are used as cofactors (Allured *et al.* 1986). The structure of DT with NAD (Bell and Eisenberg 1996), demonstrated that NAD binds DT, in a manner different to that observed for dehydrogenases, with a conformation which was presumed to aid cleavage of the N-glycosidic bond on catalysis. Additionally, DT undergoes structural change upon NAD binding, resulting in a loop near the active site (now termed the active site loop) becoming disordered (Bell and Eisenberg 1996) (Figure 2.4). The loop, which is stabilised by a hydrogen bond network in the NAD free structure, was proposed to be disrupted by NAD interference with its bulky hydrophobic residues. This disruption was speculated to be an important step prior to protein substrate binding (Bell and Eisenberg 1997).

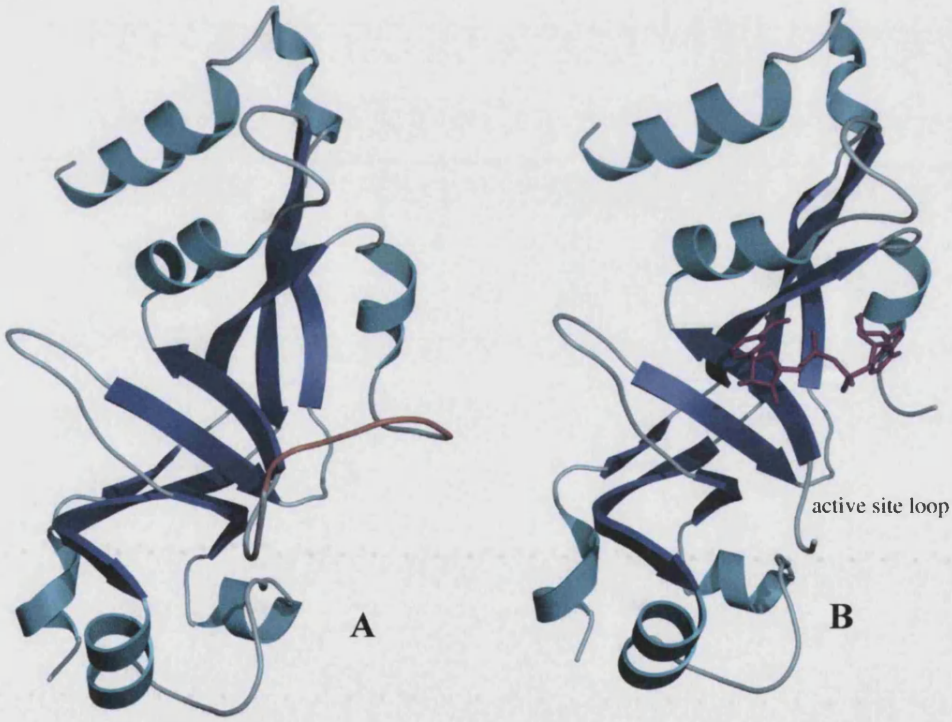


Figure 2.4: A, Diphtheria toxin (PDB=1SGK) (Bell and Eisenberg 1997) with α-helices in turquoise, β-strands in blue and the active site loop in orange. **B, Diphtheria toxin in complex with NAD** in pink (PDB=1TOX) (Bell and Eisenberg 1996) and the active site loop missing.

By comparison, the binary toxins have not been as well studied as the A:B toxins and the most well characterised are C2 from *Clostridium botulinum*, the Iota toxin from *Clostridium perfringens* and the vegetative insecticidal protein (VIP) from *Bacillus cereus*. The first extensive analysis of the residues involved in catalysis for a binary toxin, C2, confirmed that both the conserved catalytic glutamate and the second conserved glutamate in the binary toxins were both essential for ADP-ribosyltransferase activity. Also, the first serine of the conserved STX (also known as STS for binary toxins) motif and the conserved arginine in the binary and C3-like toxins (labelled 2 in Figure 2.3) were essential for this activity (Barth *et al.* 1998).

In the presence of NAD, but absence of protein substrate, all ADP-ribosyltransferases will undergo a slow NAD hydrolysis activity, effectively transferring the ADP-ribose group to water. Assays of this activity can be used in site directed mutagenesis studies to categorise residues as important for NAD binding/hydrolysis rather than protein substrate binding/transferase activity *in vitro*. Interestingly, mutation of the second conserved glutamate abolished ADP-ribosyltransferase activity but had very little effect on NAD hydrolysis suggesting that this residue is only important for the former (Barth *et al.* 1998). Mutation of the threonine of the STX motif was found to reduce the glycohydrolase activity suggesting a possible role in binding or hydrolysing NAD.

The first crystal structure of a binary toxin, VIP2 (2 denoting the enzymatic component rather than the cell binding component), and its complex with NAD, showed it to be a modular protein with a catalytic C-terminal domain and a homologous N-terminal domain which is required for interacting with the cell binding component (as shown for the Iota toxin in Figure 2.2) (Han *et al.* 1999). Both domains share a similar overall fold to the catalytic domains of DT and CT. They consist of a core comprising a five-stranded mixed β-sheet packed against a three-stranded antiparallel β-sheet flanked by five α-helices. However, they lack an active site loop equivalent to that in DT (Han *et al.* 1999). NAD bound in the cleft between the five-stranded β-sheet and the third α-helix (α8) (Figure 2.5). The flexibility of the toxin was revealed by the four independent molecules of VIP2 in the unit cell being of two distinct conformations, depending on a 6° rotation of the sub-domain including α8 towards the active site. By contrast, the NAD complex was in an intermediate conformation with a 3° rotation (Han *et al.* 1999) (Figure 2.5).

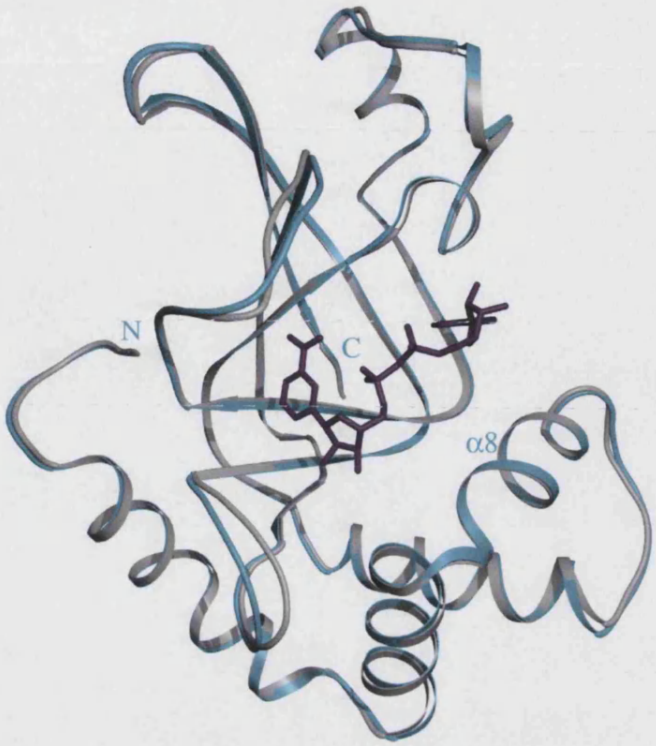


Figure 2.5: Vegetative insecticidal protein 2 (VIP2) from *B. cereus*. (PDB=1QS2) (Han *et al.* 1999). Native structure in blue and NAD complex in grey, NAD in pink. The N and C termini are labelled, along with helix α8.

Subsequently, the structural and extensive mutational analysis experiments have also been carried out for the enzymatic component of the Iota toxin (Ia). The Iota toxin shares an identical topology to VIP2 with the exception of an extra helix in the N-terminal domain (Tsuge *et al.* 2003). Complex structures with NADH and NADPH showed these substrate analogues to bind in a similar way to NAD with VIP2 (Tsuge *et al.* 2003). Mutational and kinetic analyses showed that similar residues to those identified for C2 are important for NAD binding and ADP-ribosyltransferase activity (Tsuge *et al.* 2003). Additionally, Y246, conserved in the binary and C3-like toxins (Figure 2.3), was also shown to be kinetically important for binding the protein substrate, actin. Another interesting finding was that and an asparagine on the third helix (α8), was shown to be important for protein substrate specificity (Sakurai *et al.* 2003).

2.3 Rho-GTPases structure and regulation

C3 from *Clostridium botulinum* has been widely used as a tool to study the Rho GTPases due to its specificity for RhoA, B and C and not other Rho GTPases such as Rac and Cdc42. RhoA-C have 83% sequence identity whereas the Rho GTPase family, including Rac and Cdc42, only share 55% sequence identity as shown by the sequence alignment in Figure 2.6. The Rho GTPases are part of a larger family of Ras-related GTPases which have 30% sequence identity between them. The Ras and Rho families of GTPases play a central role in many cell signalling pathways and are regulated through the binding and hydrolysis of GTP. There are many members of the Rho GTPase family and these include RhoA-E, RhoG, RhoH, Rac 1-3, Cdc42, TC10, TTF and Rnd1-3. Of these, RhoA, Rac1 and Cdc42 are the best characterised and have been widely studied for their roles in organisation of the actin cytoskeleton.

The Rho GTPases cycle between the GTP bound active form and the GDP bound inactive form of the enzyme as shown in Figure 2.7. GDP is exchanged for GTP through the action of guanosine nucleotide exchange factors, or GEFs, leading to a change in conformation allowing interactions with downstream effector molecules. Rho activation occurs at the membrane, where it is anchored via one or two farnesyl or geranylgeranyl groups attached to a C-terminal cysteine containing motif. The signal is terminated by GTPase activating proteins, or GAPs, which enhance the low intrinsic rate of GTPase hydrolysis leaving the GTPase bound to GDP. The system is further regulated by GDP dissociation inhibitors, or GDIs, which bind the inactive GTPase-GDP complex, preventing association with GTP which is more abundant in the cell. The GDIs bind the membrane anchor and thereby relocalise the GTPase in the cytosol away from its effector proteins.

The structures of Rho and the related Ras GTPases have been extensively studied (for recent review see (Paduch *et al.* 2001)) and have revealed two switch regions which alter shape depending on whether GDP or GTP are bound (Figure 2.8). Interactions of the GTPase with signalling effector molecules therefore depends of the conformation of the switch regions. Binding of GDP requires coordination by Mg^{2+} which binds the β-phosphate oxygen (Wei *et al.* 1997). The magnesium itself is coordinated by three

Figure 2.6: Sequence alignment of some Rho family proteins and RalA. Conserved residues for all seven proteins are shown in red. The two switch regions are highlighted in yellow and the phosphate-binding loop in blue. Residues shown to be involved in binding C3bot1 in RhoA are highlighted in grey where conserved. Rac I33 and RhoA T37 (red) and their equivalents are shown in bold, as is the Rho catalytic residue, Q63. Alignment done with T-Coffee (Notredame *et al*. 2000).

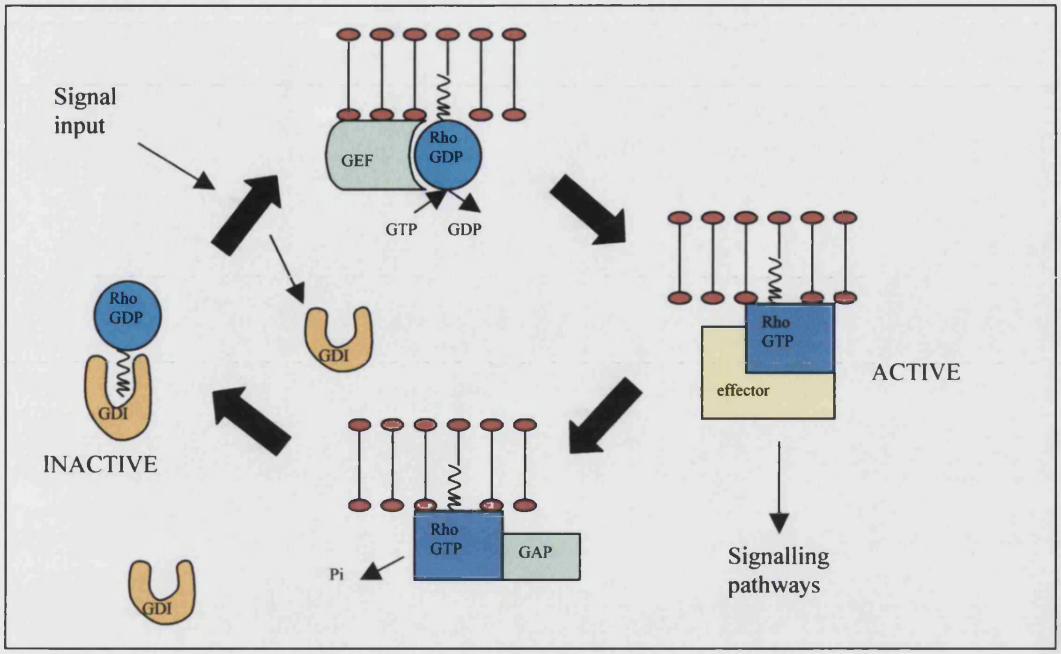


Figure 2.7: Flow diagram of GTPase activation. Rho is activated at the membrane via exchange of GDP for GTP by guanine exchange factors (GEFs). Rho can then signal to effectors at the membrane. Rho is inactivated via hydrolysis of GTP, aided by GTPase activating proteins (GAPs). Rho can then be further down-regulated by translocation to the cytosol via binding of GDP dissociation inhibitors (GDIs).

water molecules and the carbonyl oxygen of Thr 37 in switch I. The interaction of GEFs with switch I was hypothesised to dislodge the Mg^{2+}, and therefore GDP, through movement of Thr 37 (Wei *et al.* 1997). This was confirmed by the structure of RhoA-GDP without Mg^{2+} which shows a large movement of switch I away from the GDP binding site leaving it in an open conformation (Shimizu *et al.* 2000). The binding of GTP also requires Mg^{2+} but in the structure of RhoA with the non-hydrolysable GTP analogue, GTPγS, the magnesium was shown to be shifted 1Å from its position in the GDP structure. Also, it is coordinated, similar to as seen in the Ras-GTPγS structure, between the β and γ phosphates with hydrogen bonds from Thr 19, Thr 37 and two waters (Ihara *et al.* 1998) (This is illustrated later in Figure 2.9). The GTPγS structure also showed the expected large rearrangements in the switch regions compared to the GDP bound form (Figure 2.8).

Hydrophobic residues in switch I flip out to shield the GTP from hydrolysis, but also help stabilise the conformation of switch II which holds the catalytic residue, Gln 63 (Ihara *et al.* 1998).

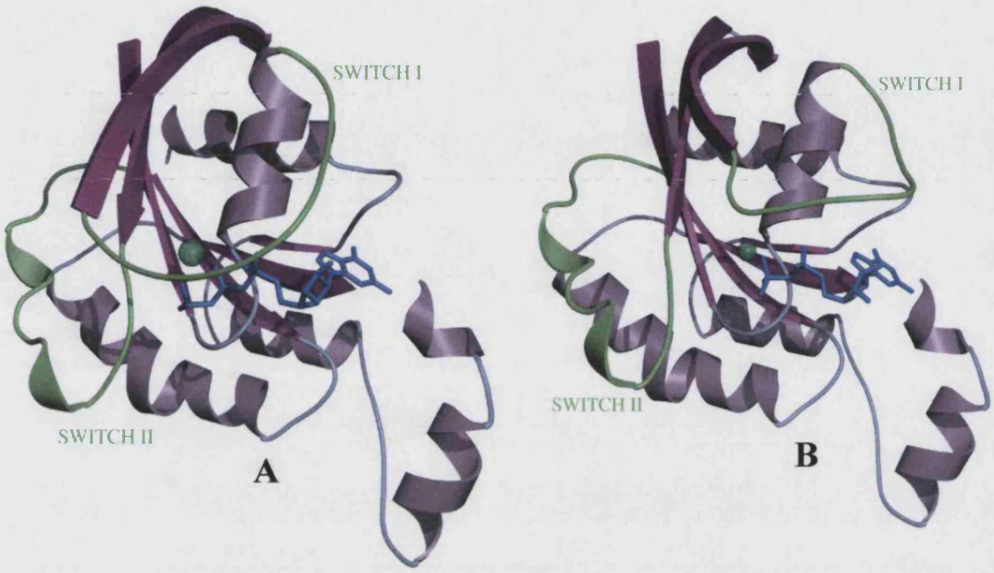


Figure 2.8: A, RhoA in complex with a GTP analogue (**A**) (PDB=1A2B) (Ihara *et al.* 1998) **and GDP(B)** (PDB=1FTN) (Wei *et al.* 1997) in blue, highlighting the switch regions and magnesium in green.

The structures of GAPs and GEFs and their complexes with Rho GTPases, as well as the use of transition state analogues, have also helped clarify the mechanism of regulation. GDP and AlF_4^- /AlF_3 combinations have been very important as transition state mimics for the GTPase reaction and have been used in various Rho structures (Rittinger *et al.* 1997b; Nassar *et al.* 1998). GTP hydrolysis proceeds via an in-line nucleophilic attack on the γ phosphate by a water molecule. The transition state of the γ phosphate is planar leading to an inversion of conformation after the addition of the water. Both aluminium fluoride structures mimic the planar γ phosphate but AlF_3 mimics the trigonal arrangement of the intermediate phosphate better, whereas the AlF_4^- has the more appropriate charge. Recently, a structure has been solved using MgF_3^- which fulfils both of these requirements (Figure 2.9) (Graham *et al.* 2002).

RhoGAPs inactivate Rho by enhancing its natural GTPase activity to hydrolyse GTP to GDP. Consequently, there are many rhoGAPs and the GAP domain is often part of a larger protein with several domains for interacting with other effectors. Although there are not many conserved residues amongst this diverse range of GAPs, the few there are allowed the Rho-interface domain to be identified in the first GAP structure –p50rhoGAP (rhoGAP), and Arg 85, termed the arginine finger, hypothesised as involved in enhancing GTPase activity (Barret *et al.* 1992). The identity of the arginine finger as a transition state stabiliser was apparent from the Rho-related Cdc42Hs -p50rhoGAP –GMPPNP complex

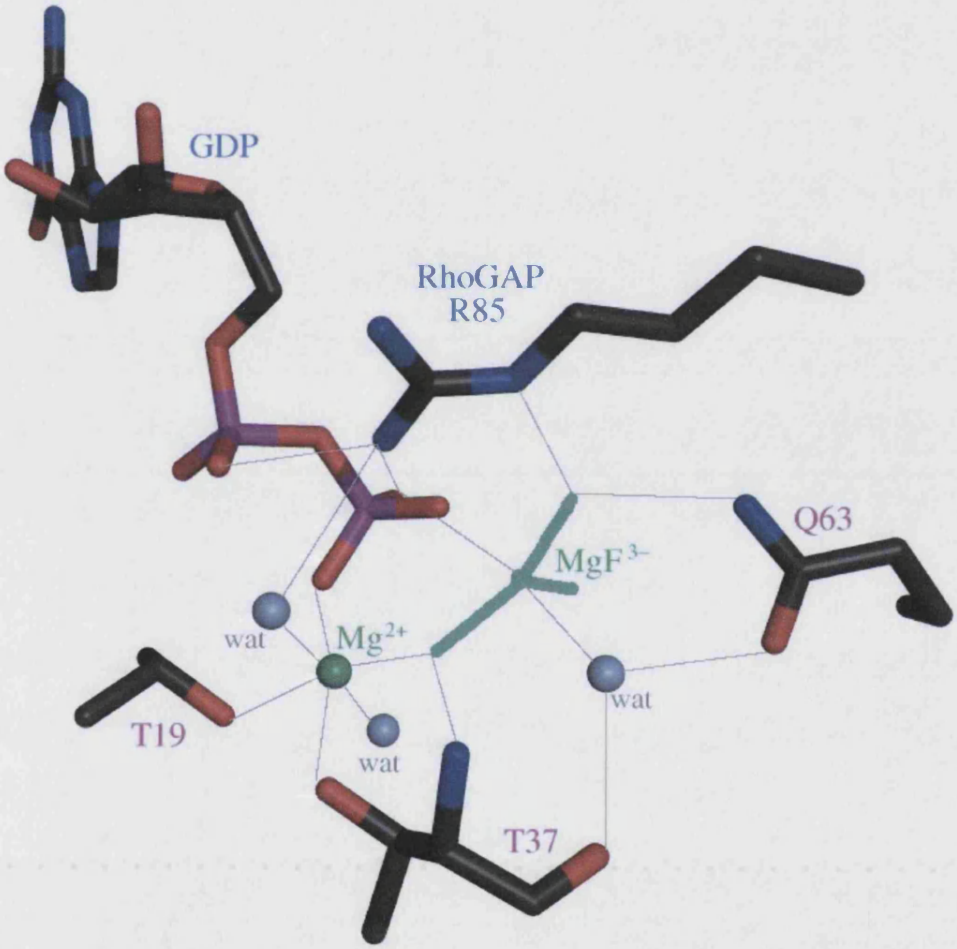


Figure 2.9: Active site of the RhoA:rhoGAP complex with a transition state analogue of GDP with MgF_3^- (PDB=10W3) (Graham *et al.* 2002). RhoA residues are labelled in pink and rhoGAP residues in blue. Bonding interactions are shown as grey lines. The planar, trigonal MgF_3^- mimics the third phosphate of GTP which is cleaved. Residues and waters coordinating to the Mg^{2+} and MgF_3^- ions are shown. The interactions of rhoGAP R85, the arginine finger, are also shown.

structure (Rittinger *et al.* 1997a). The high resolution structure of RhoA –rhoGAP-GDP.AlF_4^- confirmed the role for the arginine finger, showing that it contributed directly to the Rho active site in a position to stabilise the transition state (Rittinger *et al.* 1997b) (Figure 2.9). Additionally, recent molecular dynamics simulations of the related Ras GTPase suggest that the arginine finger may help activate the water molecule for nucleophilic attack aiding the well established catalytic Gln 61 in that role (Resat *et al.* 2001).

The arginine finger is not the only required feature for GAP activity as mutation of this residue to lysine, decreases but does not abolish the catalysis (Rittinger *et al.* 1997a). These results emphasised that the stabilisation of the two switch regions by a pocket on rhoGAP, between the B and F helices, is also important for catalysis (Figure 2.10) (Rittinger *et al.* 1997a). This was later confirmed by the structure of an alanine mutant of the arginine finger of rhoGAP in complex with Cdc42 and GDP.AlF_3 (Nassar *et al.* 1998). In this case, the arginine deficient rhoGAP was still able to stabilise the transition state. It was thought that the ordering of the two switch regions, which are highly mobile before

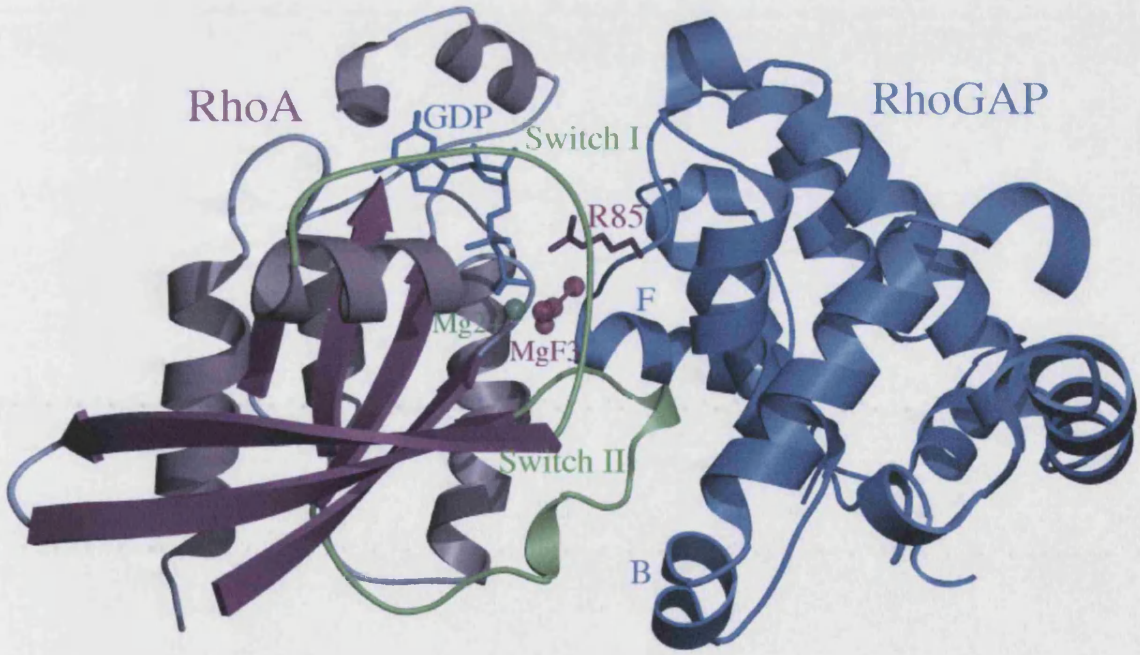


Figure 2.10: RhoA (pink) in complex with rhoGAP (blue) with GDP (blue) and MgF_3 (pink) (PDB=1OW3) (Graham *et al.* 2002). The switch regions and the Mg^{2+} are shown in green and R85, the arginine finger from rhoGAP, in pink. RhoGAP helices B and F are labelled.

rhoGAP binding, is key. Indeed, mutational analyses have identified two asparagines, and to a lesser extent, a second arginine as essential for stabilising switches I and II respectively (Fidyk and Cerione 2002). The triple mutant is virtually inactive with a 500-fold decrease in activity, indicating that these residues are the most critical in catalysis.

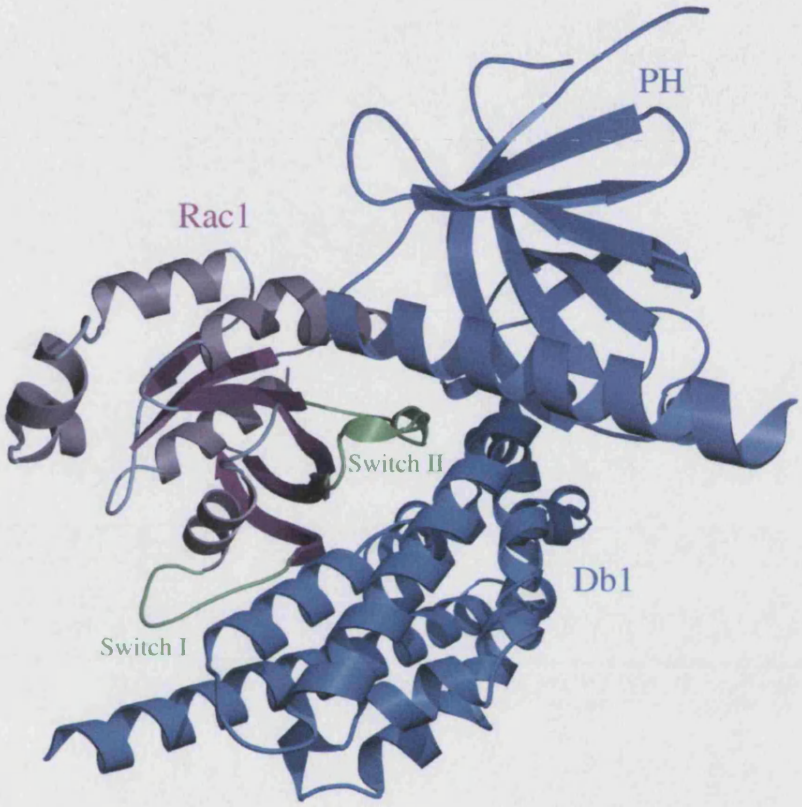


Figure 2.11: Rac1 (pink) in complex with the Db1 and PH domains (blue) of Tiam1 (PDB=1LB1) (Snyder *et al.* 2002). The switch domains are shown in green.

Like GAPs, the GEF domain is diverse and is often part of a larger protein with other modules for different signalling pathways. The most common type of GEFs for Rho GTPases are the Dbl-homology, or DH, domains. These domains are always found in partnership with pleckstrin-homology, or PH domains, which are thought to possibly play a role in the localisation or regulation of DH. Three structures of Rho proteins with DH-PH domains have been solved, Rac1 with Tiam1 (Worthylake *et al.* 2000), and Cdc42 and RhoA with Dbs (Snyder *et al.* 2002).

The structure of Rac1-Tiam1 (Figure 2.11) showed that binding of DH to Rac1 caused a rearrangement of switch I of which the two major consequences are movement of Thr 35 and Ile 33. Thr 35 (equivalent to RhoA Thr 37) was moved away from the Mg^{2+} binding site, preventing it coordinating the ion, whereas Ile 33 (equivalent to Val 35 in RhoA -Figure 2.6) is moved towards the ribose binding site, sterically hindering the binding of the GDP. Switch II also shows conformational change moving Ala 59 towards the Mg^{2+} binding site preventing magnesium binding. This positioning of the alanine is supported by a GEF lysine which is highly conserved amongst GEFs. Interestingly, the complex of Rac1 with the bacterial toxin SopE, which acts as a GEF, shows that SopE holds Rac1 in the same conformation as Tiam1 despite a very different structure (Buchwald *et al.* 2002). It appears to act with the same mechanism, but using different amino acids.

The Rac-Tiam1 complex showed no interaction of the PH domain with Rac1 suggesting that its role in ensuring full GEF activity is stabilisation of the DH domain (Worthylake *et al.* 2000). However, in the complex of Cdc42 with Dbs, the PH has a more direct role in stabilising DH residues involved in Cdc42 binding and actually interacts with Cdc42 itself (Rossman *et al.* 2002). Surprisingly, mutation of the PH-binding Cdc42 residues shows no decrease in the activity *in vitro* (Rossman *et al.* 2002).

Unlike the GAPs and the GEFs, there is a lot less diversity amongst the GDIs. There are two groups which have no significant sequence homology: the rabGDI's and the rhoGDI's, although the members within these two families are similar. RhoGDI's are about 200 residues in length and regulate all Rho family proteins. The NMR structure of rhoGDI revealed that there are two domains. The N terminal domain is unstructured in solution and comprises the first third of the protein, whereas the C-terminal domain which has an immunoglobulin-like fold, comprises the last two thirds of the protein (Keep *et al.* 1997). The crystal structure of the C-terminal domain (Keep *et al.* 1997) identified a hydrophobic cavity which was hypothesised as the binding site for the Rho GTPase geranylgeranyl membrane anchor. This was confirmed by NMR experiments, showing that isoprenyl groups interact with the structured domain in solution, and the crystal structure of Rac1-rhoGDI, where the binding of its geranylgeranyl group was observed (Grizot *et al.* 2001). The crystal structures of the Rac2-rhoGDI (Scheffzek *et al.* 2000) and RhoA-

rhoGDI (Longenecker *et al.* 1999) complexes showed that the N-terminal domain of the rhoGDI does become ordered on binding Rho GTPases (Figure 2.12). The core of the N-terminal domain forms a helical hairpin which stacks against and stabilises the switch II. RhoGDI also contacts switch I and appears to prevent GDP dissociation by stabilisation of the Mg^{2+} without actually blocking the nucleotide binding site (Scheffzek *et al.* 2000). The interaction with the switch regions also inhibits the binding of GAPs or GEFs.

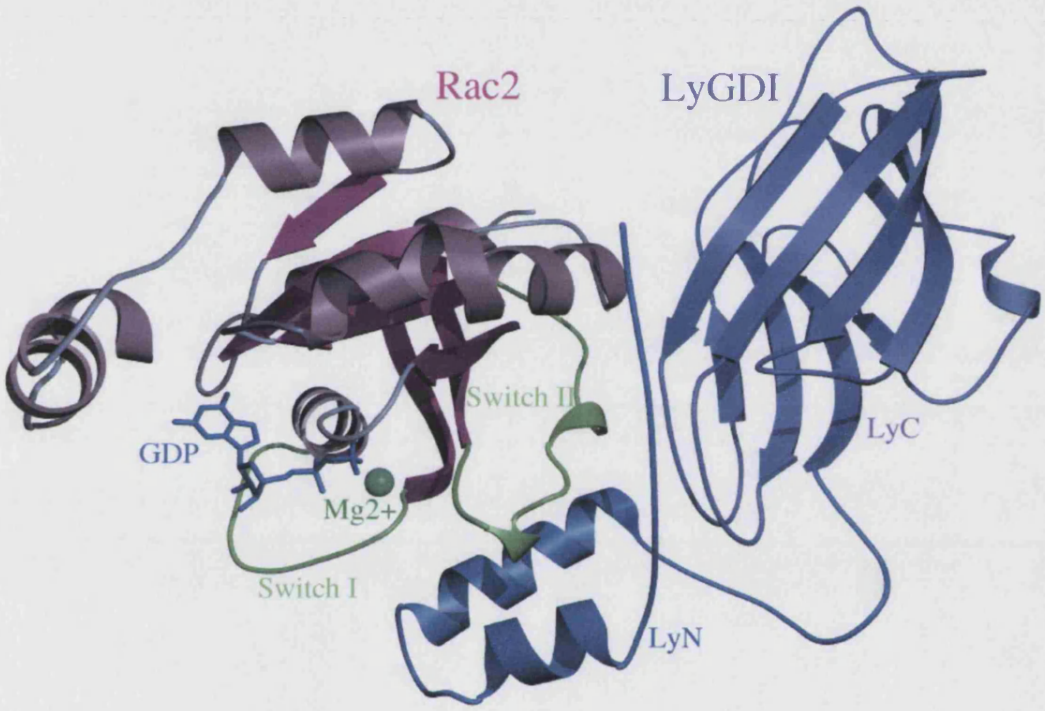


Figure 2.12: Rac2 (pink) in complex with LyGDI (blue) (PDB=1CC0) (Scheffzek *et al.* 2000). The switch regions and Mg^{2+} are shown in green and GDP in blue. The N-terminal domain of GDI (LyN) becomes ordered on binding.

2.4 The history of C3 from Clostridium botulinum and C3 exoenzymes

The C3 exoenzyme from *Clostridium botulinum* was initially discovered during studies of the botulinum neurotoxins. It was noted that preparations of neurotoxins C and D also appeared to have ADP-ribosyltransferase activity associated with them (Ohashi and Narumiya 1987; Didsbury *et al.* 1989). However, it was later shown that this was due to contamination with C3 (Aktories and Frevert 1987a; Rosener *et al.* 1987; Adam-Vizi *et al.* 1988). The confusion was probably due to the expression of C3 concurrently with the neurotoxin C1 on the same phage (Rubin *et al.* 1988). Therefore, the C3 exoenzyme from *C. botulinum* was originally named as it was the third toxin to be found in strain C of *C. botulinum* after the neurotoxin (C1) and the ADP-ribosyltransferase binary toxin C2. C3 was first identified in 1987 as being distinct from C2 as it is half the size, at about 24kDa, and has a cellular protein target with a mass of around 20kDa, instead of actin (Aktories *et al.* 1987b). The nature of the substrate for C3 was only known by its size, 20-22kDa, the same substrate that was incorrectly reported for botulinum neurotoxin D (Ohashi and Narumiya 1987). Following research showed it to be a GTP-binding protein, as its ribosylation was inhibited by GTP[S] (Aktories and Frevert 1987a). Also, it was thought to be related to the ras family of G-proteins as when injected into cells it produced similar, but not identical, morphological changes. This research later pioneered the way for studying Rho GTPase signalling. It showed that when introduced into cells by osmotic shock, C3 produces distinctive cell rounding in NIH 3T3 fibroblasts or growth of neurites in neuronal PC12 cells (Rubin *et al.* 1988; Morii and Narumiya 1995).

Surprisingly, the first observation of the C3 ribosylation target as the small G-protein Rho, was incorrectly attributed to the enzymatic activity of botulinum neurotoxin C1 (Kikuchi *et al.* 1988). The paper also noted that the enzyme ribosylated three very similar substrates, one of which was identified as 'Rho', and that the ribosylation was dependent on Mg^{2+}. The ribosylation reaction was characterised in more detail and confirmed as attributable to C3 after more thorough purification of the enzymatic agent (Aktories *et al.* 1988). Amino acid sequencing of the purified substrate confirmed it to be Rho (Narumiya *et al.* 1988; Braun *et al.* 1989). Further experiments showed that recombinant RhoA (Aktories *et al.* 1989) and RhoC (Chardin *et al.* 1989), the product of

two of the three Rho genes known of at that time, were substrates for C3 *in vitro*. Experiments with RhoC also demonstrated that the reversible morphological changes observed in Vero cells, on addition of C3, were due to dissolution of actin filaments as shown by fluorescent staining of F-actin (Chardin *et al.* 1989). These experiments suggested a probable role for Rho in actin filament regulation.

A more detailed study of the effect of a constitutively active Rho, Val14Rho, and C3 ribosylated Rho in Swiss 3T3 cells (Paterson *et al.* 1990) agreed with this role for Rho in actin filament organisation. Injection of ribosylated Rho caused cells to round up with the same morphology as C3 treated cells, even though endogenous Rho was present, whereas ribosylated Val14Rho had no effect on cells. This suggested that ribosylated normal Rho in its GDP-bound form competes with the endogenous Rho for GEFs, but could not function normally due to the ribosylation (Sehr *et al.* 1998). Also, the ribosylated, constitutively active Val14Rho in its GTP-bound form would not be able to bind GEFs and so not interfere with the endogenous Rho. During the meanwhile*, a more detailed amino acid analysis of the ribosylated Rho identified an asparagine residue, later known as Asn 41, as the target residue (Sekine *et al.* 1989). It was also shown that injection of the constitutively active, ribosylation resistant Val14/Ile41Rho prevented the C3 induced rounding up of cells (Paterson *et al.* 1990). However, injection of ribosylation resistant normal Rho could not rescue the C3 induced phenotype.

These experiments give some insight into the complexity of Rho signalling and it is now known that the three C3 targets, Rho types A-C (Aktories *et al.* 1989; Chardin *et al.* 1989; Williamson *et al.* 1990) (Figure 2.6), vary in their functions which may have contributed to this confusion. Also, much research has shown that the localisation of signalling molecules, such as Rho, in the cell can be of primary importance to their function. In normal cell signalling in human fibroblasts, activation of RhoA leads to recruitment to specific areas of the plasma membrane enriched with actin, where it binds its effectors. Treatment with C3 has been suggested to cause mislocalisation to other membrane areas via interference with the localisation signal, preventing the signalling (Michaely *et al.* 1999). Also, scrape loading of C3 into rat fibroblasts has been shown to actually lead to degradation of RhoA (Malcom *et al.* 1996). Alternatively, viral expression

* Alternative to meanwhile or during the meantime from Monty Python's Flying Circus

of C3 in human adenocarcinoma cells showed that not all the RhoA in the cells was accessible to C3, probably depending on location (Meacci *et al.* 1999). Recently, it has been suggested that ribosylation of RhoA causes it to be trapped in the GDI complex (Genth *et al.* 2003).

The cloning of two C3 genes from *Clostridium botulinum* in 1991 allowed easier study of the enzyme circumventing the necessity to work with lethal bacteria and botulinum neurotoxins. The latter, however, could be precipitated out of solution heating the impure protein to 100°C for five minutes, taking advantage of C3s thermostability (Aktories *et al.* 1988). The C3 clones from type C strain 468 and type D strain 1873 are identical (Popoff *et al.* 1991) and were later classified (Wilde and Aktories 2001) as C3bot1. The clone from type C strain 003-9 (Nemoto *et al.* 1991) was later known as C3bot2. Both C3bots are synthesised with a 40 residue propeptide which is cleaved to produce the mature enzyme with a molecular mass of 23KDa (Morii and Narumiya 1995), and a calculated pI of approximately 9.8 (Nemoto *et al.* 1991).

Further studies added insight into the details of the ADP-ribosylation reaction. It was confirmed that GDP-bound Rho was a better substrate for C3bot than Rho-GTP and that confusions over this is the past were probably due to the different affinities of Rho for GDP and GTP (Habermann *et al.* 1991). EDTA therefore inhibits ribosylation as it sequesters the Mg^{2+} ions essential for Rho nucleotide binding (Habermann *et al.* 1991). It was also noted that although the forward ribosylation reaction had a pH optimum of 7.5, the reverse reaction, creating NAD from nicotinamide and ADP-ribose could be observed optimally at pH 5.5 (Habermann *et al.* 1991).

Fortunately, information about the C3 active site residues proved to be more conclusive than the mechanism of Rho inhibition. Site directed mutagenesis cleared up the confusion from two reports. The first was used [^{32}P]nicotinamide-2-azidoadenine dinucleotide as a photoaffinity probe in C3bot to identify residue 9-19, in particular Trp 18, as part of the binding site (Chavan *et al.* 1992). The second used [carbonyl-^{14}C] NAD with UV irradiation of the homologous C3 from *Clostridium limosum*, which shares 63% sequence identity with C3bot (Just *et al.* 1992), to identify Glu 174 as the catalytic residue (Jung *et al.* 1993). It was shown that mutants with substitutions of Trp 18 remained fully active whilst any mutation of Glu 174 caused a large decrease in activity, indicating the

central role of Glu 174 (also known as Glu 214 if not numbered from the start of the mature protein) in catalysis (Saito *et al.* 1995).

Despite the ever increasing characterisation of C3, any evidence as to a possible role for C3 in *Clostridium botulinum* pathology was missing. The lack of binding or translocation domain, only allowing cell entry by pinocytosis, and the high toxicity of the botulinum neurotoxins with which C3 is secreted (Rubin *et al.* 1988) does not suggest an obvious role. However, this has not prevented C3 from becoming a widely used tool in Rho research, due to its specificity for RhoA-C and not for the related proteins Rac and Cdc42 (Aktories *et al.* 1989; Chardin *et al.* 1989; Just *et al.* 1992). This specificity was investigated through protein engineering of the Rho and Rac proteins (Wilde *et al.* 2000). Different domains of RhoA were substituted for the equivalent domain in Rac to identify the first ninety residues as necessary for ribosylation to occur. Secondly, site directed mutagenesis of Rac with RhoA residues identified six amino acids, Arg 5, Lys 6, Glu 40, Val 43, Glu 47 and Glu 54 (Figure 2.6) as important for the formation of the C3-RhoA complex.

C3 has also been shown to interact with the ras related RalA GTPase. Interestingly, RalA is not ribosylated by C3 but appears to be an inhibitor of ribosylation (Wilde *et al.* 2002a). Of the Ras-related GTPases, RalA is most similar to RasH, N and K with 48% sequence identity (Chardin and Tavitian 1986) and is thought to be involved in numerous signalling cascades including regulation of the cytoskeleton (Ohta *et al.* 1999), vesicle trafficking and endocytosis (Nakashima *et al.* 1999; Jullien-Flores *et al.* 2000; Moskalenko *et al.* 2002; Sugihara *et al.* 2002). The affinity of RalA for the homologous C3 from *C. limosum*, is very high with a K_d of 12nM (Wilde *et al.* 2002a). Also, RalA does not appear to bind to C3 at the same site as RhoA, but appears to increase the NAD-glycohydrolase activity of C3. This suggests that it binds C3 in a way that optimises the glycohydrolase activity but blocks the binding of RhoA. Although in some cases, this additional activity of C3 may have had a complex effect on experiments designed to characterise RhoA, this is likely to be quite small as the interaction is not enzymatic.

The ability of C3 to stimulate neurite outgrowth in PC12 cells (Rubin *et al.* 1988) has led to further research showing that C3, through inactivation of Rho, can also stimulate axon regeneration *in vitro* and overcome growth inhibition *in vivo* (Lehmann *et al.* 1999).

It has been proposed that C3 has potential as a therapeutic for stimulating differentiation of neurones (Watanabe *et al.* 2000). As a development of this C3 has been engineered for easier cell entry. In one report, several short peptides, the most effective of which was a short proline rich peptide, were added to its carboxyl-terminus to increase the uptake of C3 into neuronal cells, whilst maintaining its activity (Winton *et al.* 2002). C3 has also been engineered to include a fragment of the N-terminal domain of the *Clostridium perfringens* Iota toxin Ia. This allows the translocation of Ia into the cell, with the help for the cell binding component, Ib, by receptor mediated endocytosis. It was found that a 129 residue stretch of Ia was sufficient to transfer active C3 into Vero cells in an Ib dependent manner (Marvaud *et al.* 2002).

The structure of C3bot1 from *C. botulinum* was solved in 2001 showing it to consist of a mixed α/β fold with a β-sandwich core (Han *et al.* 2001) (Figure 2.13). A three stranded anti-parallel β-sheet flanked by the first four helices is stacked against a five stranded mixed β-sheet flanked by helix five. The topology was identical to the catalytic domain of the binary toxin, VIP2 from *Bacillus cereus* (Han *et al.* 1999), which suggested a similar NAD binding site positioned between the β-sheet core and helix α3. The related DT-like toxins have an active site loop in the position of α3 of C3bot1, but comparison with VIP2 suggested a unique active site for the C3-like and binary toxins. This corresponds to residues 207-214 in C3bot1, of which Glu 214 (also known as Glu 174) is the conserved catalytic glutamate, named the ADP-ribosylating toxin turn-turn (ARTT) motif. It was hypothesised that that Gln 212, the conserved asparagine in C3-like exoenzymes, could have a role in binding the target residue of Rho, Asn 41. However, it has since been shown by site-directed mutagenesis that it is not involved in NAD or Rho binding, but is essential for stabilising the C3bot1-NAD-Rho complex for ribosylation (Menetrey *et al.* 2002). Phe 209, which also resides in this loop, is thought to be involved in binding to a hydrophobic patch on the Rho surface near Asn 41 (Han *et al.* 2001).

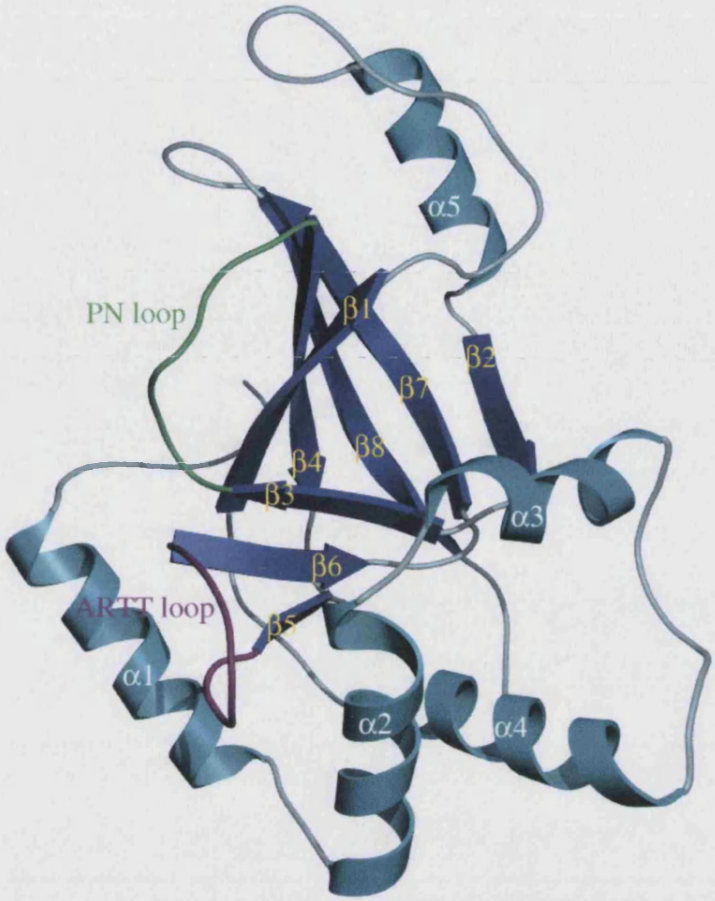


Figure 2.13: C3bot1 from *C. botulinum* (PDB=1G24) (Han *et al.* 2001). The ARTT loop is shown in pink and the PN loop in green. Helices and strands are labelled.

The crystal structure of C3bot1 in complex with NAD gave further insight into the residues involved in catalysis (Menetrey *et al.* 2002). C3bot1 undergoes conformational changes upon binding NAD including the ARTT loop and the NAD binding loop (between strands β3-β4), termed the phosphate-nicotinamide (PN) loop. It also undergoes a 'crab-claw' movement of a large sub-domain, comprising β2, β7, β8 and α5, which closes the active site around the NAD, similar to that seen for VIP2. This report also confirmed by site-directed mutagenesis that the conserved 'STS' motif, which had been speculated to be important for active site stability as it is in the binary toxins (Han *et al.* 2001), was not essential for NAD binding or ribosylation.

The family of C3-like exoenzymes is rapidly growing as more proteins with sequence similarity to C3bot1 are identified. The other C3-like exoenzymes characterised so far, apart from C3bot1 and C3bot2, include C3lim from *Clostridium limosum*, C3cer from *Bacillus cereus*, and three enzymes from *Staphylococcus aureus* known as EDIN A (C3stau1), C3stau2 (EDIN B) and C3stau3 (EDIN C). Early characterisation of C3lim included the calculation of the K_m for NAD to be 0.3μM and the specific ADP-ribosyltransferase activity to be 3.1 nmol/mg platelet membrane protein/min (Just *et al.* 1992). Photolabelling of C3lim with [*carbonyl*-^{14}C] NAD also highlighted the catalytic site to be within the sequence equivalent to 211-219 in C3bot1 (Jung *et al.* 1993), suggesting that the conserved glutamate (Glu 214) was also important for catalysis in C3lim.

C3cer has recently been cloned and analysed by site directed mutagenesis based on previous structures and analyses and shown to have the same functional residues as C3bot1 and C3stau2 (Wilde *et al.* 2003). Additionally, mutation of a threonine, the first residue in the ARTT loop (serine in C3bot1), caused a large decrease in the transferase activity with little effect on the NAD glycohydrolase activity (Wilde *et al.* 2003). However, the precise role of this residue in catalysis is unclear at present. Interestingly, it was noted that its specific activity for RhoB was lower than for RhoA and even less for RhoC. This pattern also appeared to be mirrored by other C3 exoenzymes including C3lim, C3bot1 and C3stau2. Of the C3 exoenzymes from *S. aureus*, C3stau2 is the best characterised an the only one to have its structure solved (Evans *et al.* 2003). These enzymes will be discussed more fully in chapter four.

2.5 Botulinum neurotoxin structure and mechanism

The Botulinum neurotoxins (BoNTs) are produced by *Clostridium botulinum* in seven distinct serotypes, BoNT A-G, and are responsible for the deadly food poisoning known as botulism. Botulism is characterised by flaccid muscle paralysis due to the inhibition of acetylcholine release from neuromuscular junctions. The deadly tetanus toxin from *Clostridium tetani* is also part of the same family of neurotoxins. Its mode of action is to cause spastic paralysis, however, by inhibition of neurotransmitters such as glycine and GABA from inhibitory neurones in the spinal cord. These neurotoxins all share over 30% sequence identity with each other, and in particular, the catalytic domains of tetanus toxin and BoNT/B share 51.6% sequence identity. Both the botulinum and the tetanus toxins are synthesised as a 150kDa polypeptide which is post-translationally cleaved to yield a 50kDa light chain and a 100kDa heavy chain. The light chain, which holds the toxin enzymatic

Figure 2.14: Botulinum neurotoxin A, (PDB=3BTA) (Lacy *et al.* 1998) showing the binding domain in green, the translocation domain in pink and the endopeptidase domain in blue with the catalytic zinc ion in grey.

activity, cleaves intracellular proteins to block transmitter release. The heavy chain consists of both the translocation domain at its N-terminus, and the receptor binding domain at its C-terminus. The two chains remain connected by a disulphide bond which is essential for cell entry but which must be cleaved for release of the light chain into the cytosol for toxic activity.

The crystal structures of both BoNT/A (Lacy *et al.* 1998), and B (Hanson and Stevens 2000; Swaminathan and Eswaramoorthy 2000) share the same topology and have a clear domain structure (Figure 2.14). The binding domain consists of two equivalent sized β-strand sub-domains connected by an α-helix. The N-terminal sub-domain comprises a jelly roll motif whereas the C-terminal one forms a modified β-trefoil fold. The binding domain tilts away from the translocation domain leaving all surfaces open for binding. The central part of the translocation domain consists of two long, anti-parallel, amphipathic, kinked helices wrapped around each other, two helical bundles and a disordered area that was previously suggested to be a putative membrane spanning helix by the program TMAP (Persson and Argos 1996). The catalytic domain is globular and contains the zinc binding motif HEXXH, comprising the endopeptidase active site. The other main feature of the toxin is the belt region of the translocation domain which surrounds the catalytic domain and partly blocks the active site.

Intoxication by botulinum neurotoxins usually results from toxin contaminated food, or occasionally by a food derived infection of *Clostridium botulinum* in the gut in young children (Simpson 2000). The neurotoxins are carried by the blood to their target cells at the neuromuscular junction where they are taken up via receptor mediated endocytosis (Dolly *et al.* 1984). The binding mechanism is thought to involve two receptors; a low affinity receptor comprising a ganglioside which then facilitates binding to a high affinity protein receptor which allows toxin internalisation (Montecucco 1986). Antibodies against these gangliosides have shown that they are essential to BoNT binding (Kozaki *et al.* 1998). The evidence for the identity of the protein receptors varies, but the best characterised is for N-terminal region of synaptotagmin as the receptor for BoNT/B, to which it bound in the presence of gangliosides GT1b and GT1a (Nishiki *et al.* 1994; Nishiki *et al.* 1996; Kozaki *et al.* 1998) (Figure 2.15).

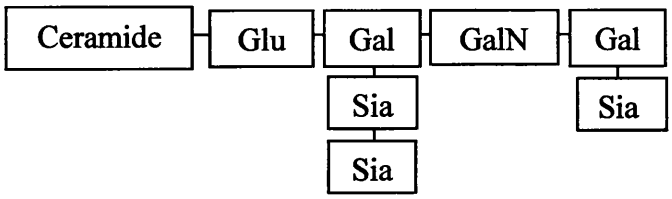


Figure 2.15: Schematic diagram of ganglioside GT1b. Gangliosides comprise a sialic acid containing oligosaccharide attached to ceramide. BoNTs require gangliosides with two sialic acid groups attached to the internal galactose for binding. Glu is glucose, Gal, galactose, GalN, galactosamine and Sia, sialic acid.

Although the necessity for ganglioside binding has been well established, the details of the binding site and the mechanism of the binding are not well understood. The ganglioside binding site has been best characterised for tetanus toxin, where the binding domain (H_C) has been crystallised with an analogue of GT1b, minus the ceramide (Fontinou *et al.* 2001) (Figure 2.16). Interestingly, this structure indicated binding sites on the toxin for both the galactose/galactosamine moiety and the internal double sialic acid moiety (Figure 2.15), causing the adjacent domains in the crystal to be cross-linked (Fontinou *et al.* 2001). This suggested that cross-linking of several toxins may be a possible mechanism *in vivo*. More recent toxin-ganglioside binding studies show that this is not the case but speculate that the sialic acid binding site may play a part in the difference in specificity between tetanus toxin and the BoNTs (Rummel *et al.* 2003). The only BoNT structure with a ganglioside mimic is the BoNT/B-sialyllactose complex (Swaminathan and Eswaramoorthy 2000), where the sialic acid is observed binding at a similar site to that seen for the galactose/galactosamine moiety in tetanus toxin (Figure 2.16).

Once the toxin has been internalised through receptor-mediated endocytosis, the catalytic light chain needs to be released for activity. For the catalytic light chain to reach the cytosol, it must translocate through the vesicle membrane for which it requires the low pH of the endosome (Simpson *et al.* 1994). The translocation domain forms cation channels in the membrane at low pH (Hoch *et al.* 1985; Blaustein *et al.* 1987; Shone *et al.* 1987) and has recently been shown to act as a channel to translocate the light chain across

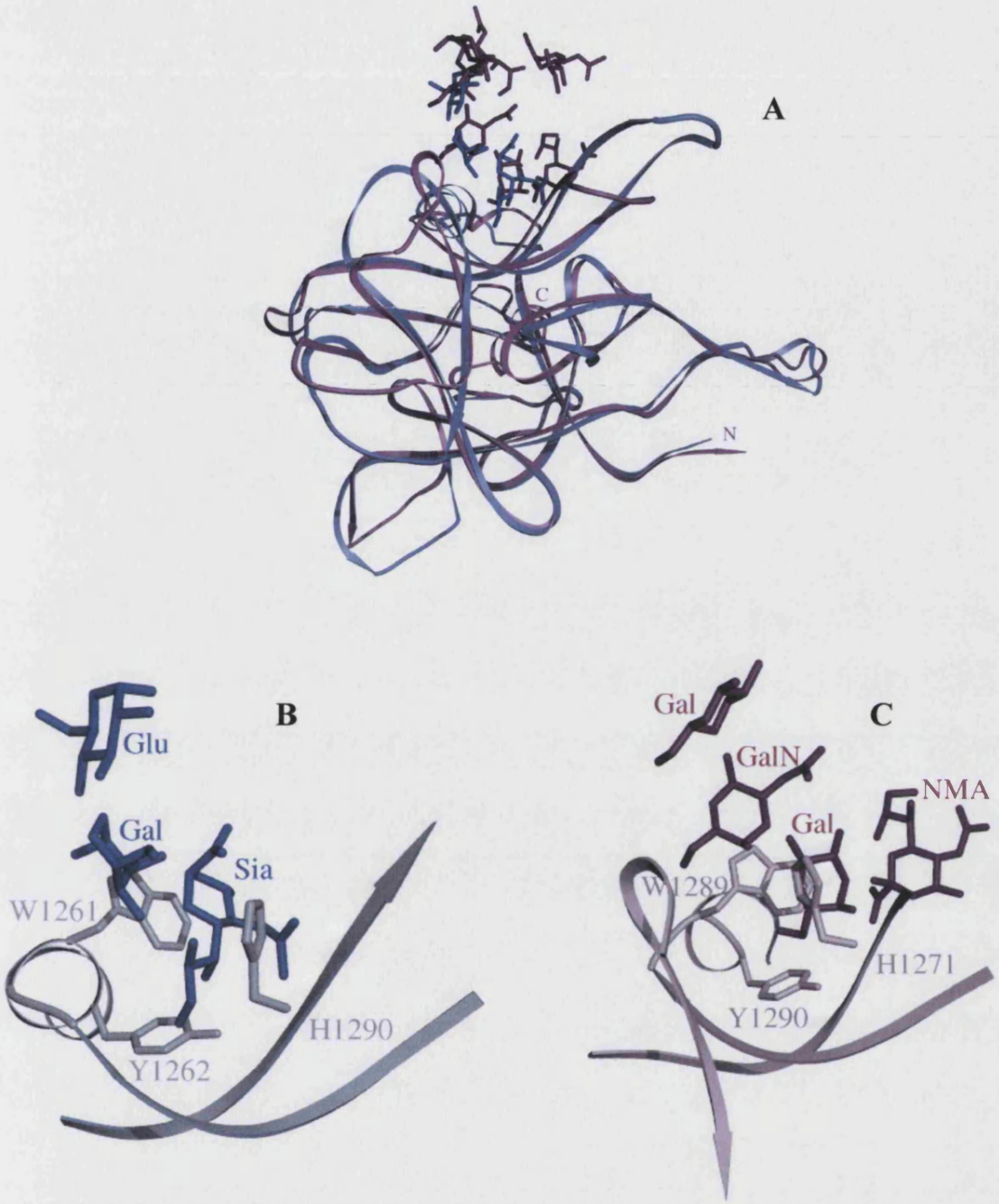


Figure 2.16: Binding of sugars to tetanus toxin and BoNT/B. A: The β-trefoil domain of tetanus toxin (PDB=1FV3) (Fontinou *et al.* 2001) and BoNT/B (PDB=1F31) (Swaminathan and Eswaramoorthy 2000) in complex with a GT1b analogue and sialyllactose respectively. BoNT/B is shown in blue and tetanus toxin in pink. **B:** Close up of the sialyllactose binding of BoNT/B. Monomers of the sialyllactose are labelled showing that the sialic acid is located in the binding cleft. **C:** Close up part of the GT1b analogue binding to tetanus toxin. Monomers of the GT1b analogue are labelled. The galactose moiety is centrally located in the cleft, equivalent to the position of sialic acid for BoNT/B.

the membrane (Koriazova and Montal 2002). Using circular dichroism, with channel currents and the light chain protease activity as assays, it was shown that at pH 5, the light chain partially unfolds and blocks the HC channel under non-reducing conditions. However, under reducing conditions, the light chain was translocated across the membrane and retrieves its catalytic activity (Koriazova and Montal 2002). The proposed mechanism for translocation is that the HC undergoes structural change and inserts in the membrane at the low endosome pH and the light chain partially unfolds. The light chain is then translocated through the channel formed by the HC, refolds, possibly through a chaperone–like activity of the HC, and dissociates from the HC in the reducing conditions of the cytosol.

Once inside the cell, the neurotoxins all perform the same zinc dependent endopeptidase reaction. The target of the protease action is the SNARE family which facilitate exocytosis of the neurotransmitter. Three SNARE proteins are required for correct exocytosis and the different neurotoxins cleave one of the three. BoNT/B, BoNT/D, BoNT/F and BoNT/G all cleave synaptobrevin/VAMP, whereas BoNT/A and BoNT/E cleave SNAP25 and BoNT/C cleaves syntaxin (for recent review see (Turton *et al.* 2002)). The fact that between them, the botulinum toxins cleave all three proteins involved in membrane docking, has allowed them to be used as a tool to study the roles of these proteins in transmitter release.

Chapter 3

The Structure of C3bot1

3.1 Introduction

At the outset, the aim of this project was to crystallise and determine the structure of C3bot1 in order to characterise C3 exoenzymes and to help elucidate the basis for their protein substrate specificity. Whilst purification of C3bot1 was still underway, the structure was solved elsewhere (Han *et al.* 2001), but this project was continued in order to crystallise a complex with one of the substrates. Crystallisation attempts followed the published ones for the original structure and soaks were carried out with NAD, nicotinamide, ADP-ribose and 3-acetylpyridine adenine dinucleotide, a substrate analogue. Whilst these experiments were going on, the structure of C3bot1 in complex with NAD was also solved elsewhere (Menetrey *et al.* 2002). Unfortunately, none of our structures showed any binding of the substrates. However, of the two crystal forms described in the original paper (Han *et al.* 2001), the second one, with which we had the most success, could not be refined by the authors. Therefore, this chapter describes the purification, crystallisation and structure determination of C3bot1 in this crystal form and analyses the differences in crystal packing.

3.2 Methods and Results

3.2.1 Purification of C3bot1

All protein for purification was provided as *E. coli* cell paste by Dr. J.M. Sutton at the Health Protection Agency (HPA), Porton Down. Although purification of recombinant C3bot1 has been described previously, this purification procedure resulted in a yield of only 3mg C3bot1/litre cells (Morii and Narumiya 1995). Therefore, C3bot1 was initially expressed as a His-tagged protein in order to aid purification. The addition of a His tag, however, resulted in the majority of the recombinant protein being located in the insoluble fraction upon lysis by sonication. Refolding of this protein was attempted by solubilisation in 7M guanidine HCl followed by drop wise addition into a variety of buffers, 50mM Tris/HCl pH 7.5 or pH 8.5 plus 0.1M NaCl with additives 1% PEG 400, 20% glycerol, 33mM Chaps or 1M guanidine HCl or 50mM BisTris pH 6.5, 50mM Sodium phosphate and 0.5M arginine. Refolding in these buffers proved unsuccessful. Therefore, an alternative approach of purifying the protein in 8M urea was tried as the binding of the His

tag to the metal affinity column does not require a correctly folded protein. However, only a small percentage of the denatured protein bound to cobalt or nickel affinity columns.

Following the lack of success with the His-tagged protein, a C3bot1-MBP construct was created by Dr J.M. Sutton at the HPA. MBP tags are often used to improve the solubility of proteins but have the disadvantage of large size and the necessity of cleavage during purification (Sachdev and Chirgwin 2000). The *Apa*I-*Hind*III fragment from the expression vector pMALc2x (New England Biolabs) was isolated and subcloned into the vector pBC SK+ (Stratagene) to generate the vector pBCmalE. The construction of the vector was confirmed by restriction digest and sequencing. A *Eco*RI fragment containing the C3bot1 cloning sequence was subcloned into pBCmalE digested *Eco*RI and the clone verified. The majority of the recombinantly expressed C3bot1-MBP was soluble upon cell lysis by sonication but only a maximum of 2.5mg fusion protein (a small percentage of that expressed as assessed by SDS-PAGE) bound the amylose affinity column (in 20mM Tris pH 7.5) from half a litre of cell culture. To improve this yield, an initial step involving purification by SP-sepharose column in 20mM Na.Hepes pH7.3 was included. A large peak was eluted from the SP-sepharose column by an ascending NaCl gradient. However, this only improved the yield after amylose affinity chromatography to 5mgs per half a litre of cell culture.

As the yield from the amylose column was so low, this step was omitted and the C3bot1-MBP purified by SP-sepharose chromatography was cleaved by Factor Xa (Novagen) overnight at room temperature as C3bot1 is stable at high temperatures. The enzyme concentration used was optimised to 0.5U Factor Xa (Novagen) per mg fusion protein. This concentration was required for complete cleavage as C3bot1-MBP was not easily purified from C3bot1. The C3bot1 was further purified away from the MBP by SP-sepharose chromatography in a buffer of 20mM Na.Hepes pH 7.3, 20mM NaCl. The MBP did not bind the column and the C3bot1 was eluted on an ascending NaCl gradient. If further purification was necessary, NAD affinity chromatography was performed with a Pharmacia HiTrap Blue Sepharose column (Amersham) (Figure 3.1). A final purification step by gel filtration was added which also allowed buffer exchange into 20mM Hepes pH7.3, 20mM NaCl, 2mM $MgCl_2$, 1mM EDTA. The purified C3bot1 proved to be very soluble and was easily concentrated to 39mg/ml for storage at -70°C. The yield from this purification was approximately 10mg per litre cells. The final protocol is outlined below.

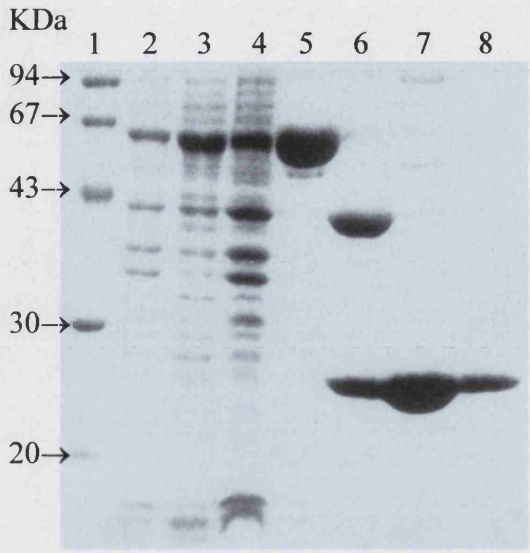


Figure 3.1: 12% SDS-PAGE gel of the C3bot1 purification. Lane1, Low molecular weight marker (Amersham); lane2, cell extract; lane 3, soluble extract; lane 4, insoluble cell extract; lane 5, peak from 1st SP-Sepharose column; lane 6, cleaved protein; lane 7, peak from 2nd SP-Sepharose column; lane 8, peak from the HiTrap Blue Sepharose column (Amersham). C3bot1-MBP, MBP and C3bot1 have molecular weights of 67KDa, 43KDa and 24KDa respectively.

For expression the clone TB1 pBCmalE C3bot1 was grown overnight at 30°C in Terrific Broth supplemented with 35 µg/ml chloramphenicol and 0.5% (w/v) glucose. The overnight culture was diluted 1:5 in fresh media and grown for 4 hours. The culture was induced with IPTG at a final concentration of 500µM and grown for a further 2 hours at 25°C before harvest. The cells were collected by centrifugation and stored at -20°C. Cells, from 1 litre of culture, resuspended in 30mls lysis buffer (20mM Hepes pH 7.3, 20mM NaCl, 1mM EDTA, 1mM PMSF), were lysed with 30µl 1mg/ml lysozyme on ice for an hour. The cells were lysed further by sonication, 5 bursts of 40s at 26microns (Soniprep 150), resting on ice between bursts and the cell debris were collected by centrifugation. The C3bot1-MBP was purified from the supernatant by SP-sepharose chromatography in 20mM Hepes pH 7.3 and eluted on a 0-1M NaCl gradient in two batches. Fractions

containing C3bot1-MBP were pooled and cleaved overnight at room temperature by incubation with factor Xa (Novagen), 0.5U/mg fusion protein. C3bot1 was diluted and purified further by either SP-Sepharose or Pharmacia HiTrap Blue Sepharose (Amersham) chromatography in 20mM Hepes pH 7.3 or both. The purest fractions were concentrated in a centrifugal 15KDa cut off concentrator (Amicon) and finally purified in batches of 300µl on a Superdex-200 HR 10/30 (Amersham) gel filtration column run at 0.75ml/min in 20mM Hepes pH 7.3, 20mM NaCl, 5mM $MgCl_2$ and 1mM EDTA. The pure protein was concentrated to 39mg/ml, aliquoted and stored at -70°C.

3.2.2 Crystallisation

Crystallisation was performed using the hanging drop vapour diffusion technique with precipitants based on the published conditions (Han *et al.* 2001). Crystals were grown at 19°C in a drop containing 1.5µl 35mg/ml C3bot1 in 20mM Hepes pH 7.3, 20mM NaCl, 5mM $MgCl_2$ and 1mM EDTA and 1.5µl well solution. The well solution contained 18% PEG 3350, 0.1M Imidazole/malate pH 5.5, 3% saturated sodium orthovanadate solution, 0.15M NaCl and 20mM Tris pH 7.5. Plate like crystals (Figure 3.2) that diffracted up to

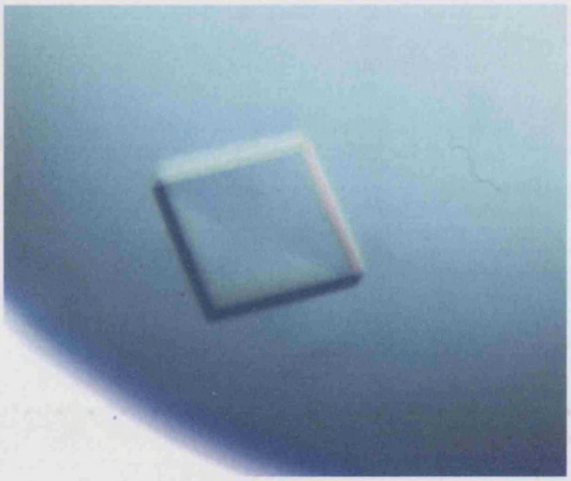


Figure 3.2: Tetragonal crystals of C3bot1 grew in two days at 19°C using the hanging drop technique at a protein concentration of 35mg/ml. 1.5µl C3bot1 in 20mM Hepes pH 7.3, 20mM NaCl, 5mM $MgCl_2$ and 1mM EDTA was mixed with a 1.5µl well solution (18% PEG 3350, 0.1M Imidazole/malate pH 5.5, 3% saturated sodium orthovanadate, 0.15M NaCl, 20mM Tris pH 7.5).

1.6Å formed within two days but were highly mosaic. Soaking these crystals in 27% PEG 3350, 0.1M Imidazole/malate pH 5.5, 3% saturated sodium orthovanadate, 0.15M NaCl and 20mM Tris pH 7.5 decreased the mosaicity in some cases.

Co-crystallisations with nicotinamide and 3-aminobenzamide were performed with the same precipitant but with either 20mM, 15mM or 10mM of the additive added. These crystals were also transferred to the higher precipitant concentration containing the corresponding level of additive. Soaks of native crystals with NAD, nicotinamide, ADP-ribose and 3-acetyl pyridine adenine dinucleotide were performed at the higher precipitant concentration containing 100mM additive.

Crystals were also grown at the other published conditions by Han *et al.* but these were even thinner, more plate-like and grew in mosaics over the whole drop making crystal manipulation difficult. Crystallisation with the later conditions published (Menetrey *et al.* 2002) was also attempted but without success.

3.2.3 Data collection

Crystals were soaked in cryoprotectant containing 27% PEG 3350, 0.1M Imidazole/malate pH 5.5, 3% saturated sodium orthovanadate solution, 0.15M NaCl and 20mM Tris pH 7.5 and 20% glycerol and cryo-cooled to 100K. One native and several derivative datasets were collected on stations PX 14.2 and PX 9.6 at the Synchrotron Radiation Source at Daresbury, UK, using an ADSC-CCD detector (Table 3.1). The highest resolution data set came from native crystals soaked in 27% PEG 3350, 0.1M Imidazole/malate pH 5.5, 3% saturated sodium orthovanadate solution, 0.15M NaCl and 20mM Tris pH 7.5 with 100mM nicotinamide for 30hrs (Figure 3.3). After refinement, no nicotinamide was observed bound so this data set was refined as the highest resolution native data set. The data were processed and scaled in the space group I4 in HKL2000 (Otwinowski and Minor 1997) to 1.89Å. The cell dimensions were a=b= 73.29, c=218.23. All the crystals were mosaic, and most highly mosaic, but the high symmetry of the space group meant that 91.6% completeness was achieved for this highest resolution data set. There were two C3bot1 monomers per asymmetric unit and the crystals contained 59.4% solvent.

Data reduction was carried out using the program TRUNCATE (CCP4 1994). The structure was determined by molecular replacement with MOLREP (Vagin and Teplyakov 1997) using the previously determined C3bot1 structure (Han *et al.* 2001). The solution for the first monomer had an R-factor of 50.2% and a correlation coefficient of 37.3, which improved to 39.7% and 61.2 respectively for both solutions.

Table 3.1: Data collection statistics for all data sets collected

Crystal type	Resolution (Å)	$\Delta\phi$ (°)	Exposure (sec)	R_{sym} (final shell) (%)[a]	Completeness (final shell) (%)
Native	2.4	0.75	10	13.1 (13.5)	92.2 (78.8)
100mM 3-aminobenzamide, 28hr soak	1.89	0.5	10	5.7 (58.4)	85.0 (70.4)
100mM Nicotinamide, 30hr soak	1.89	0.5	10	9.4 (35.7)	91.6 (91.8)
50mM ADP-ribose, 95min soak	2.35	0.5	12	4.8 (44.0)	88.6 (87.7)
20mM 3-acetylpyridine adenine dinucleotide, 20min soak	2.84	0.5	12	8.5 (47.5)	88.6 (87.9)
5mM NAD, 30min soak	2.9	0.5	12	8.5 (47.5)	94.0 (87.9)

[a] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I(h)$, where $I_i(h)$ and $I(h)$ are the ith and the mean measurements of the intensity of reflection h, respectively.

3.2.4 Refinement

The first round of rigid body and restrained refinement, carried out using Refmac5 in CCP4i (Murshudov *et al.* 1997), gave an R-factor of 29.2% and a R-free (758 reflections) of 32.9%. Inspection of initial maps showed that none of the additives were bound so the nicotinamide soak was refined as the highest resolution native data set. NCS was imposed in the first rounds of refinement and restraints were used to keep the r.m.s. bond length deviation low. Water molecules were added in the final stages of refinement by the implementation of ARP/wARP (Lamzin and Wilson 1993) in Refmac5. The water positions were checked to have heights above 3σ in the F_o-F_c map and to have a temperature factor less than ~50Å^2. The program O (Jones *et al.* 1991) was used to rebuild

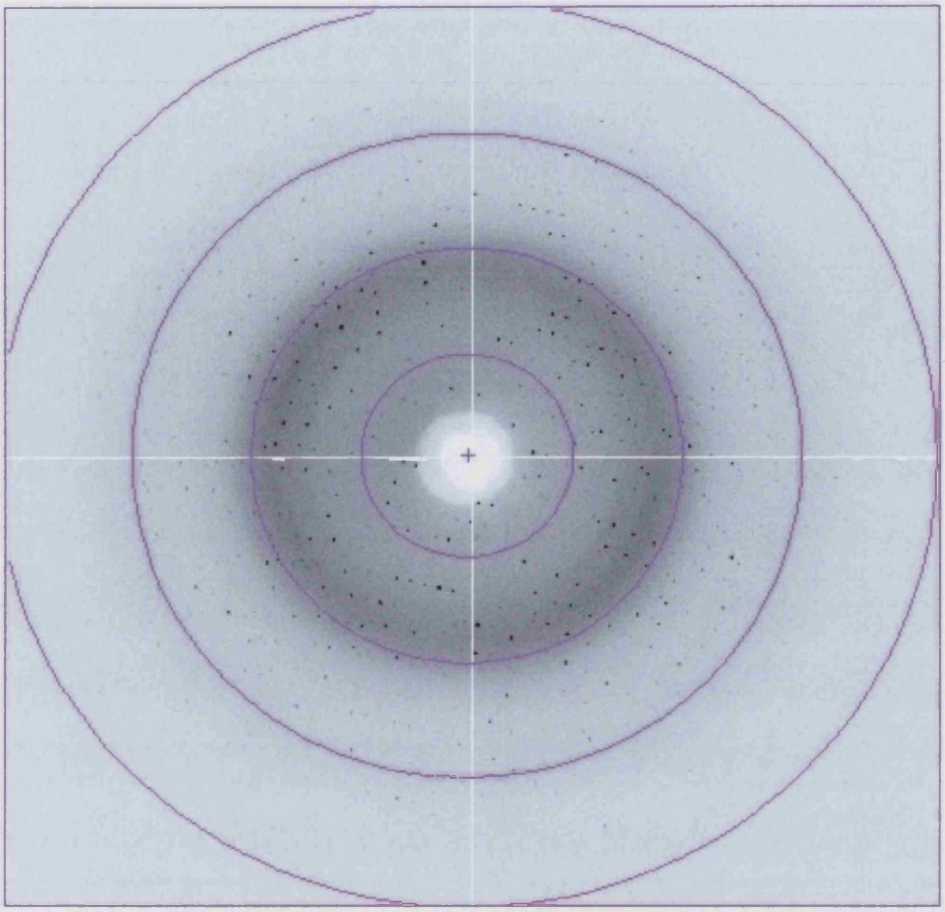


Figure 3.3: Diffraction image for C3bot1. Collected on station PX 9.6 at the SRS, Daresbury, U.K. The rings mark the resolution limits 7.64Å, 3.81Å, 2.52Å and 1.9Å proceeding outwards from the centre. The data were processed to 1.89Å.

the model and visualise the maps. The F_o-F_c map contained two peaks unexplained by the protein or solvent structure which were finally modelled as a vanadate and malate ions. The vanadate ion was modelled at 0.7 occupancy (B=30.85) to fit the height of the difference density peak. The malate ion was modelled with an occupancy of one and explained most of the difference density in that area. Its position would ideally be slightly further away from its adjacent symmetry related partner, but fits the density. The final model contains 3314 protein atoms and 327 water molecules has an R_{cryst} of 20.9% and an R_{free} value of 24.7%. Crystallographic statistics are in Table 3.2. At 1.89Å the N and C termini were clearly visible in the structure. However, the extremities of side chains of 56, 64, 65, 70, 73, 85, 94, 103, 146, 147, 157,182, 233, 235 for both molecules A and B were

Table 3.2: Crystallographic statistics for the refined data set
(30hr, 100mM nicotinamide soak)

Space group	I4
Maximum resolution (Å)	1.89
Unit cell dimensions (Å)	a=73.29 b=73.29 c=218.23
Measured reflections	683474
Unique reflections	41904
Completeness (%) (last shell)[a]	91.6 (91.8)
Mean I/σ (last shell)[a]	15.3 (4.4)
R_{sym} (%)[b] (last shell)[a]	9.4 (35.7)
R_{cryst} (%)[c]	20.9
R_{free} (%)[d]	24.7
Number of protein atoms	3314
Number of solvent atoms	327
Number of ligand atoms	
Vanadate	10
Malate	18
R.m.s. deviation in bond lengths (Å)	0.009
R.m.s. deviation in bond angles (°)	1.24
Average B factor for protein atoms (Å^2)	26.9
Average B factor for solvent molecules (Å^2)	29.9
Average B factor for ligand atoms (Å^2)	
Vanadate	30.85
Malate	44.1

[a] Last shell is 1.94-1.89Å.

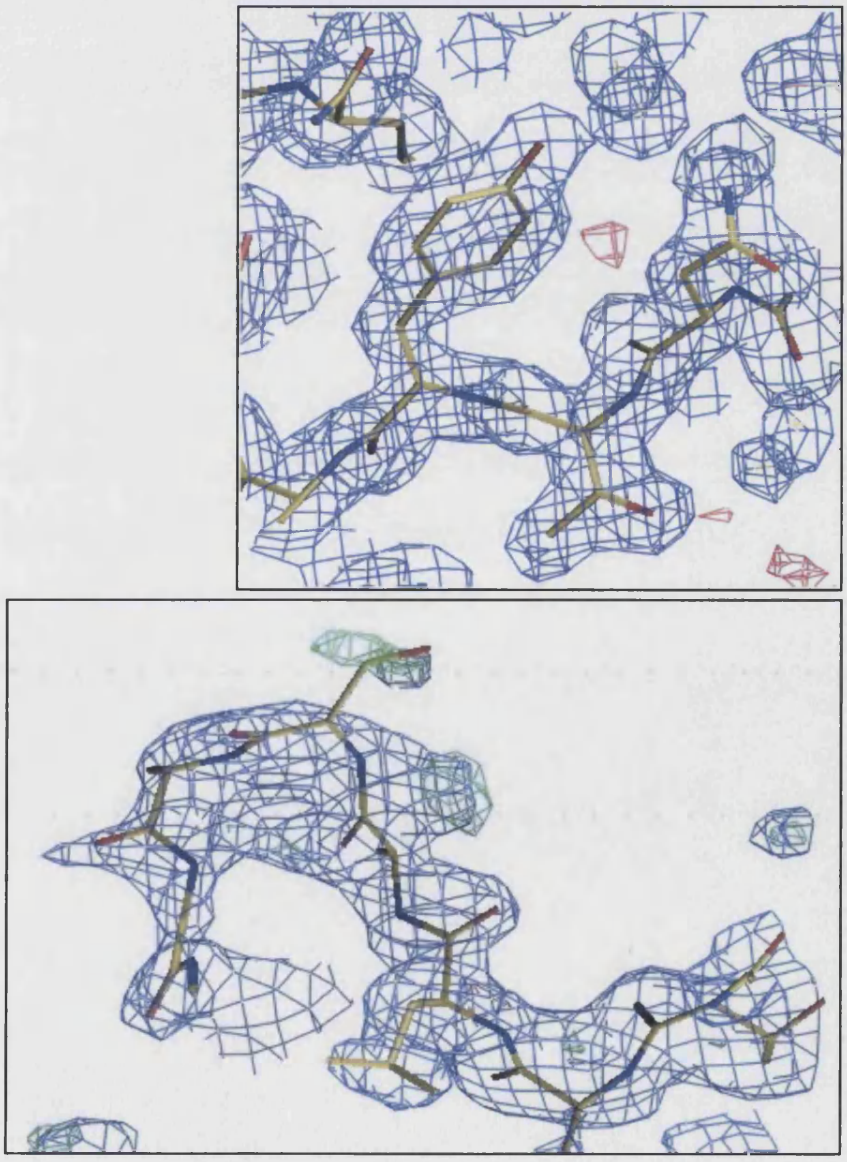
[b] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I(h)$, where $I_i(h)$ and $I(h)$ are the ith and the mean measurements of the intensity of reflection h, respectively.

[c] $R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h, respectively.

[d] R_{free} is equal to R_{sym} for a randomly-selected 758 reflections not used in the refinement (Brünger 1992).

modelled with an occupancy of zero due to insufficient density. The extremities of residues 106, 150, 153, 159, 165, 181 and 232 in molecule A and 191 in molecule B were also modelled with an occupancy of zero due to insufficient density. Figure 3.4 shows examples of the good and the poor regions of the map for the final model. The average B factor for molecules A and B were 27.0Å^2 and 26.8Å^2 respectively. The structure was checked with PROCHECK (CCP4 1994). All figure were produced with Molscript (Kraulis 1991) and rendered with Povray (www.povray.org) unless otherwise stated.

Figure 3.4: Final map and model for A, part of the N-terminus, and B, part of the variable region (residues 143-147). The C3bot1 model is shown as a stick representation, the $2F_o$-F_c map at 1σ is represented as a 3D blue mesh, the F_o-F_c map at 3σ as a green mesh (B) and as a red mesh (A). This picture was created in O (Jones *et al.* 1991).



3.2.5 Structure and crystal packing of C3bot1

The overall topology of this structure of C3bot1 (Figure 3.5) at 1.89Å in the space group I4 is identical to that of the published structures (Han *et al.* 2001; Menetrey *et al.* 2002). 94% of the residues were within the most favoured regions of the Ramachandran plot (Figure 3.6) and only two (for both A and B) were in the disallowed regions. Apart from a slightly raised B factor for molecule A, resulting in more disordered residues, there are few differences between the two molecules and they have a Cα r.m.s. deviation of 0.08Å. Even though NCS restraints were removed at the end, most side chain positions are conserved between them. The exceptions were Met 125 and Leu 143 which are both solvent exposed.

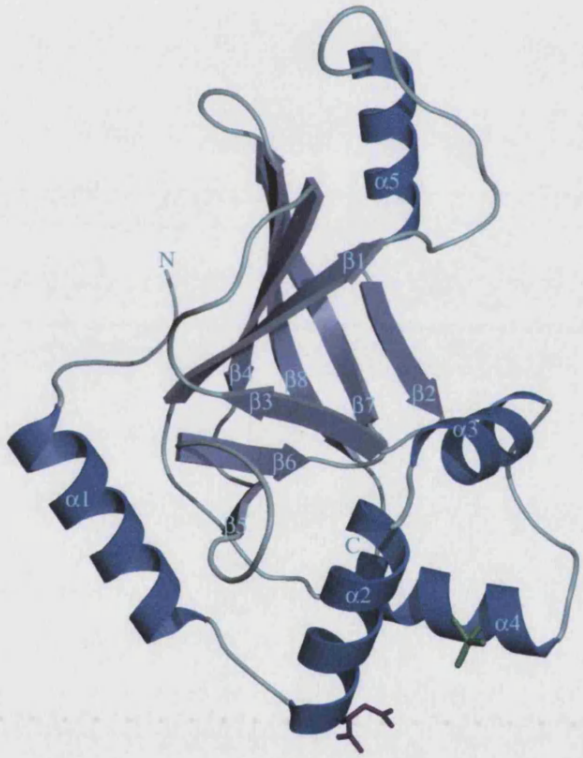


Figure 3.5: Overall structure of C3bot1 with the vanadate ion (green) and the malate ion (pink) shown. Helices are: α1, 52-66, α2, 69-81, α3, 83-92, α4, 102-114, α5, 151-162. Strands are: β1, 124-131, β2, 165-170, β3, 172-176, β4, 187-194, β5, 200-203, β6, 213-217, β7, 221-230, β8, 236-246.

In general, our C3bot1 has higher B factors than the published native structures in space group C2 (Han *et al.* 2001; Menetrey *et al.* 2002) (Figure 3.7A) which probably accounts for the larger number of disordered residues seen. These disordered residues are all solvent exposed and do not appear to play a role in catalysis. They are visible in both the published structures. We have also modelled less waters for our structure with only 163.5 on average per monomer in the asymmetric unit rather than 217.5 for the published structure (Han *et al.* 2001), which is of higher resolution. However, our C3bot1 has identical topology to this structure with a Cα r.m.s. deviation of 0.45Å for molecule A from both structures (Figure 3.7 B-D). Only residues 184-185, 147-151 and 232-236 have a different Cα backbone with Cα r.m.s. deviations of 0.01Å, 0.40Å, and 1.55Å respectively. All three regions are parts of loops which have high B factors (Figure 3.7A). Residues 147-151 are part of the region which varies most in sequence between the C3 exoenzymes (Figure 2.3, Chapter 2), termed the variable region, corresponding to helix α5

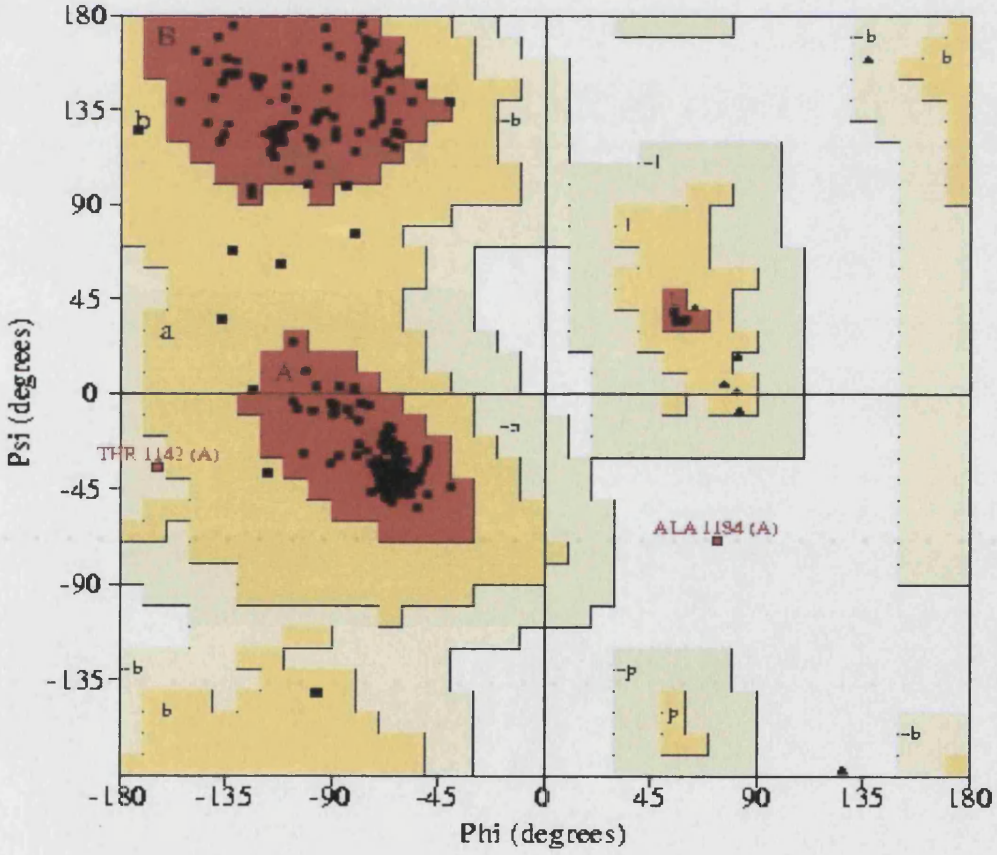


Figure 3.6: Ramachandran plot for C3bot1 molecule A.

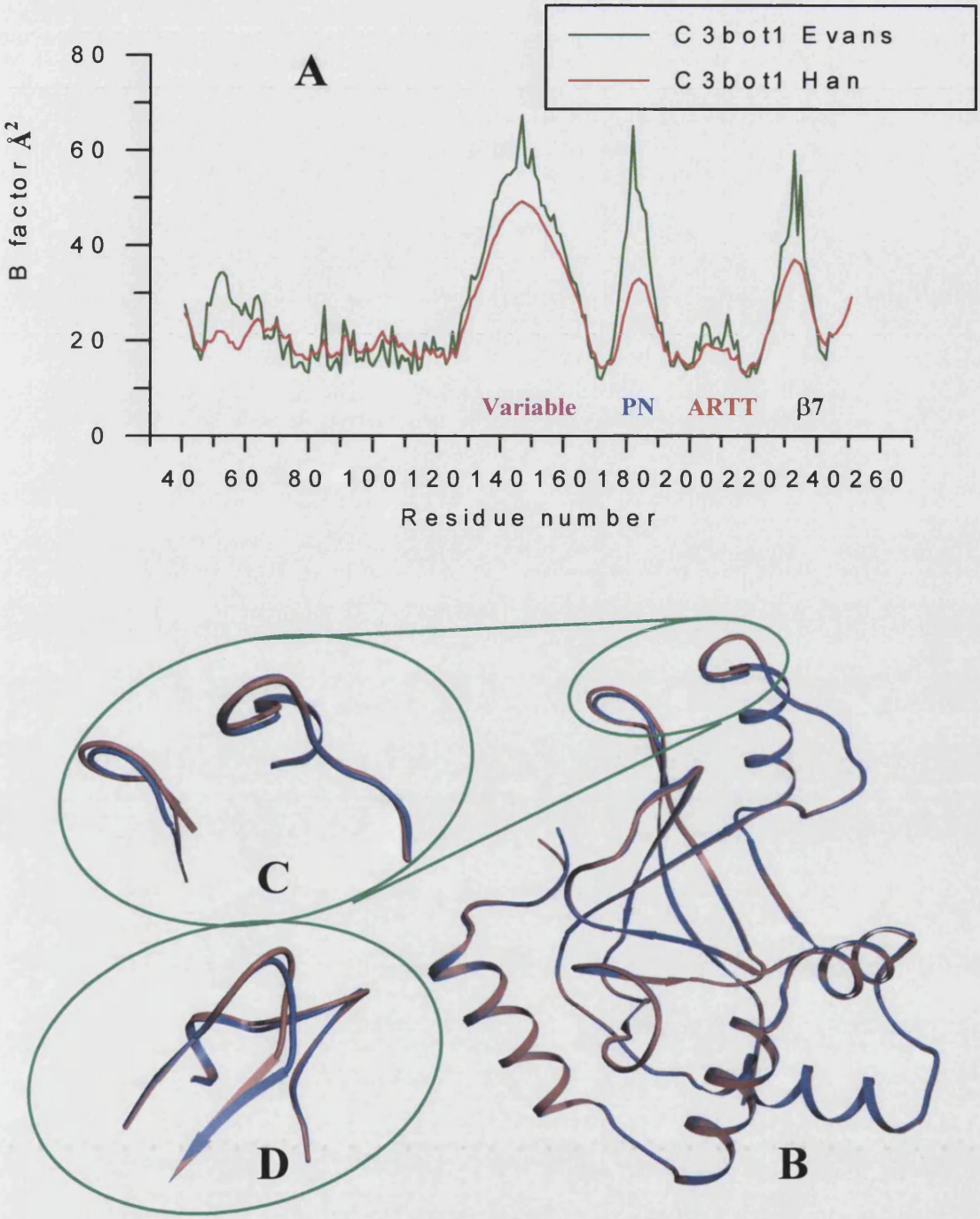
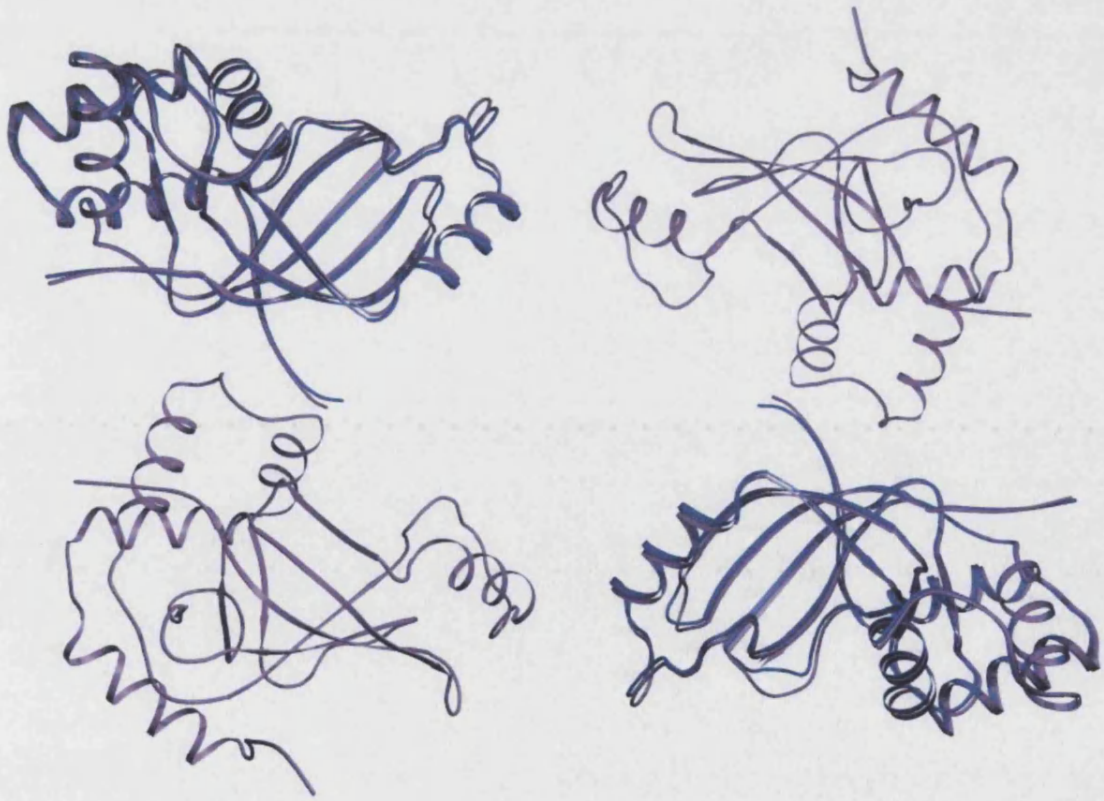


Figure 3.7: A, B factor plot for our C3bot1 and the published native C3bot1 (1G24) (Han *et al.* 2001). Approximate positions of the PN and ARTT loops, the variable region (helix α5 and the loop preceding it) and β7 are shown. **B, Superposition of our C3bot1 (purple) and the published native C3bot1 (orange). C, Insert shows closer view of the most divergent region** comprising the variable region and the loop between β7 and β8. These loops correspond to the regions in (A) that have the highest B factors. **D, A 90° rotated view of C.**

and the loop preceding it in C3bot1. Residues 232-236 are part of the loop between strands β7 and β8 which interacts with the variable region. Residues 184-185 are part of the phosphate-nicotinamide (PN) loop required for NAD binding.

Our C3bot1 is not only similar in topology to the published structures but also shares a similar packing arrangement in the asymmetric unit. Our structure has two monomers in the asymmetric unit rather than four but these two superimpose well on the one of the of the two dimers (Figure 3.8).

Figure 3.8: Superposition of our C3bot1 asymmetric unit with the published native structure (pdb=1G24) (Han *et al.* 2001). The published asymmetric unit with four molecules (pink) is superimposed on our asymmetric unit in blue.



Unlike the previous structures, however, two ions were modelled for the structure which appear to have been necessary for the crystal packing (Figure 3.5). The first, a malate ion, is located adjacent to the crystallographic axis and participates in hydrogen bonds with two symmetry related molecules as well as its own symmetry related partner (Figure 3.9). The second, a vanadate ion, is located at the intersection of three of the symmetry related molecules in the unit cell (Figure 3.10) and participates in the hydrogen bonding network.

These crystals were soaked in several substrates and substrate analogues without any binding observed. This suggested that the crystal packing may be preventing access to the active site. Analysis of the crystal packing showed this to be the case as the N-terminus of a symmetry related molecule is observed packing against helix α3 and hydrogen bonding with helix α3 and a residue in the ARTT loop (Figure 3.11). Both these regions are observed to move on binding NAD (Menetrey *et al.* 2002). It is likely that this packing arrangement not only blocks the active site, but may also prevent the conformational change required for substrate binding.

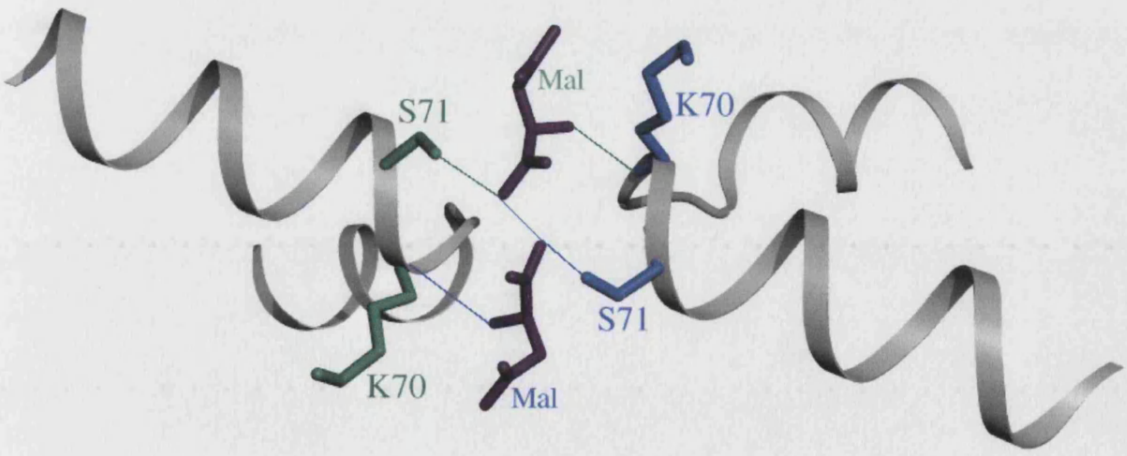


Figure 3.9: The malate ion bridges two symmetry related molecules between helices 2 and 3. Hydrogen bonds between the malates and the symmetry related molecules are coloured blue and green to match the residues.

Figure 3.10: The vanadate ion is involved in crystal packing. **A** shows two vanadate ions (pink) and four waters (pink and yellow) positioned between three adjacent molecules in the unit cell. **B** shows the involvement of the vanadate ion and three water molecules in the hydrogen bonding network in the crystal packing.

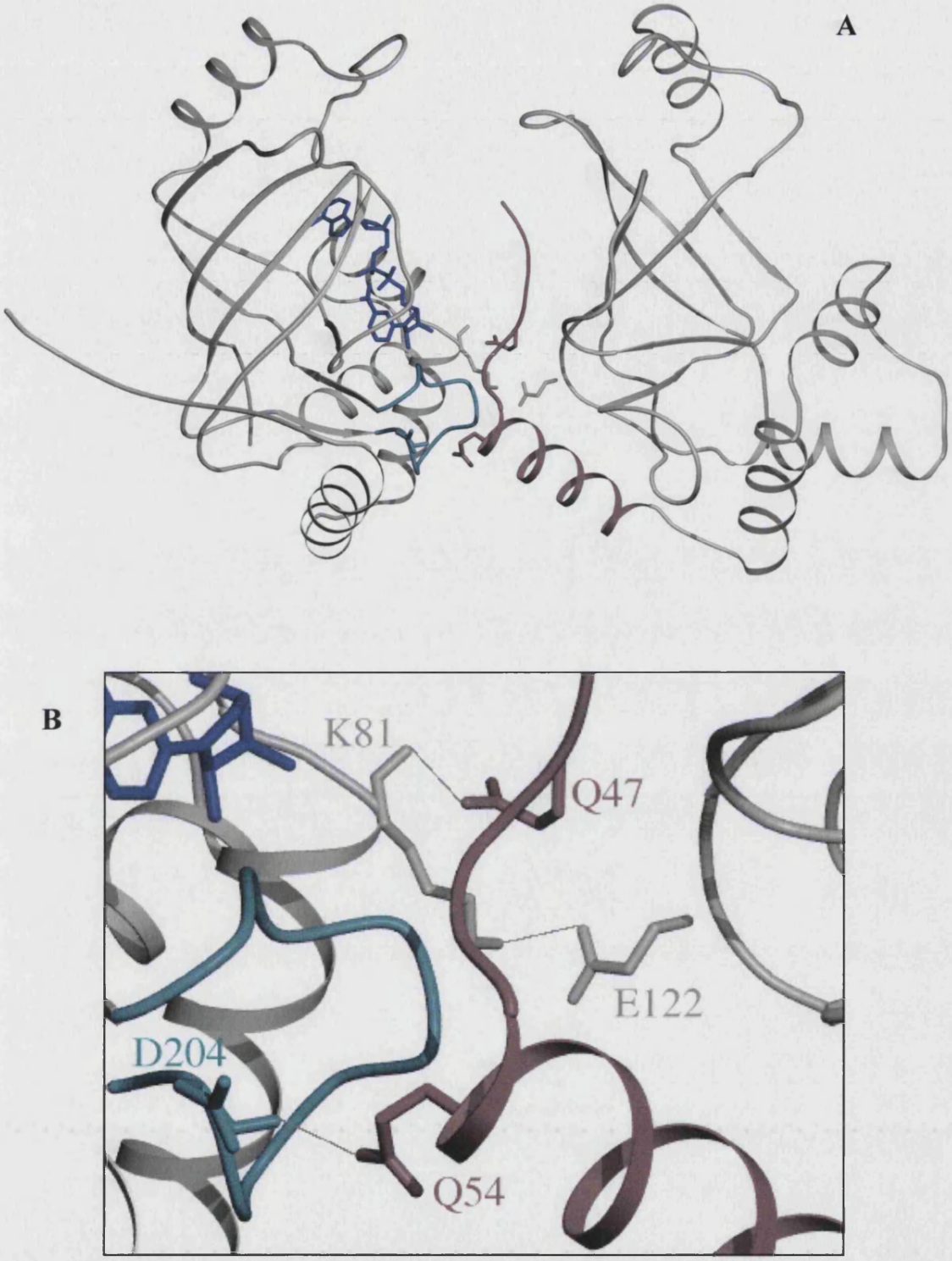


Figure 3.11: A, The N-terminus packs against helix α3 and the active site. The active site is marked by the superposition of NAD (blue) from the published crystal structure (PDB=1GZF) (Menetrey *et al.* 2002). The ARTT loop is coloured green and the N-terminus, brown. **B, Residues Q47, E122 and Q54 hydrogen bond to and electrostatically interact with residues K81 (α3) and D204 (ARTT loop) respectively.**

3.3 Discussion

Although the soluble expression and purification of C3bot1 was not straight forward at first, the final protocol utilising a C3bot1-MBP fusion protein resulted in a good yield for crystallisation. C3bot1 was also noted to be very soluble, achieving high concentrations, and reasonably thermostable as previously reported (Aktories *et al.* 1988; Han *et al.* 2001). Reproduction of the tetragonal and monoclinic crystal forms from the original paper (Han *et al.* 2001) was easily achieved using the published conditions. However, the second set of conditions published (Menetrey *et al.* 2002) never resulted in any crystals.

This data, at 1.89Å, is not as high resolution as the original publication at 1.7Å (Han *et al.* 2001) and so does not reveal much extra information about the structure, and therefore the biology, of the molecule. The lack of very complete and redundant data, maybe responsible for the messy and less well defined regions of electron density. The backbone was clearly visible throughout, however, and there were only a few side chains that were not visible at all. The regions which varied from the original publications were loop regions known to be flexible. The slight difference of position of the residues in the PN loop is not surprising as this region moves on NAD binding. The most useful observation from the structure is the packing arrangement which probably blocks the binding site. The lack of NAD binding could also be explained by the crystal pH of 5.5. NAD binds C3bot1 optimally at pH 7.5 (Habermann *et al.* 1991), although the C3bot1-NAD complex was solved at pH 3 (Menetrey *et al.* 2002). However, C3bot1 binds nicotinamide and ADP-ribose optimally at pH 5.5 (Habermann *et al.* 1991) suggesting that our crystals would have been ideal for these complexes, if access could be gained to the active site.

The previous publication of the crystallisation conditions reported diffraction to 2.15Å. This data could not be refined due to the mosaicity of the data (Han *et al.* 2001). However, our data diffracted to 1.89 Å, and although not perfect, refined well enough. There are several possibilities for these differences. The resolution difference maybe due to a shorter wavelength radiation used to collect the data (0.87Å). Alternatively our crystal quality could be better due to either the purity of the protein, any slight variation in

crystallisation conditions (e.g. fluctuations in temperature) or the soaking of our crystal in a higher precipitant concentration prior to data collection. The difference in protein purity is hard to assess, but both purifications used gel filtration as a final step. Soaking crystals in a higher precipitant concentration may have been responsible for lowering the mosaicity of our crystals to 0.57° which is much less than theirs (>1.5°). However, we found that only some of the soaked crystals showed less mosaicity than those that were not. Whatever the reason for the difference in resolution, it is possible that refinement of their data against a perfect model would be possible. The poor phases from a low homology molecular replacement model probably meant that the density was not good enough to build in the new features. The poor quality of our maps meant it was difficult to identify the extra difference density and position the malate and vanadate ions until the refinement was mostly complete. However, although the electron density is not completely interpreted in every place, it had been achieved as far as the data quality and noise of the maps would allow.

Chapter 4

The Structure of C3stau2 and its Complex with NAD

Introduction

C3stau2, from *S. aureus* is classified as part of the C3-like ADP-ribosylating exoenzyme family and shares 35% sequence identity with C3bot1 (Wilde *et al.* 2001). C3stau2, however, is more related to, and shares 78% sequence identity with, another C3-like exoenzyme from *S. aureus,* C3stau1, generally know as EDIN, or EDIN A. EDIN was first discovered in 1990 as the factor in the cultured supernatant of *S. aureus* from impetigo lesions that inhibits calcium-induced terminal differentiation in cultured mouse keratinocytes. Thus it was given the name '**E**pidermal Cell **D**ifferentiation **In**hibitor' (Sugai *et al.* 1987). EDIN was characterised as a single polypeptide with a molecular weight of 27,000 and a pI >9 (Sugai *et al.* 1990). The cloning of this protein showed the first 35 residues to be a signal sequence, which when cleaved, gave a mature protein of 212 residues which had 35% sequence identity with C3 from *C. botulinum* (C3bot1) (Inoue *et al.* 1991).

This homology with C3bot1 resulted in testing EDIN for ADP-ribosyltransferase activity. EDIN was found to ADP-ribosylate both RhoA and RhoB and a comparison of the kinetics of RhoB ribosylation with C3bot2 showed them to be similar (Sugai *et al.* 1992). EDIN and C3bot2 were also compared for their effects on keratinocytes. Although both EDIN and C3bot2 both inhibited keratinocyte differentiation *in vitro* and stimulated proliferation of human keratinocytes *in vitro* and mouse epidermis *in vivo,* EDIN has more potency for these activities (Sugai *et al.* 1992).

The role of C3s from *Clostridia* in disease has always been unclear due to no known mechanism for cell entry. By contrast, *S. aureus* has been shown to be an intracellular pathogen, offering an obvious mechanism for C3-like exoenzymes to play a role in pathogenesis. Both C3bot1 and EDIN have been shown to ribosylate RhoA-C in human umbilical cord vein cells (HUVECs) and inhibit endothelial cell migration and wound repair through the inactivation of RhoA (Aepfelbacher *et al.* 1997). EDIN has also been reported as the probable cause of acquired subglottic stenosis (ASGS) in methicillin resistant *S. aureus* (MRSA) infections of the trachea in neonates requiring intubation (Yamada *et al.* 2001). The ASGS was thought to be caused by EDIN mediated delayed wound repair of the subglottis. Prevalence of EDIN-producing strains of *S. aureus* has not

been widely studied but 12% of the clinical isolates in Japan have been reported to produce EDIN (Yamada *et al.* 2001).

C3stau2 was initially discovered in 2001 from screening clinical isolates of *S. aureus* for C3-like activity (Wilde *et al.* 2001). It was later found to be encoded on a novel pathogenicity island on the chromosome in tandem with exfoliative toxin D, ETD, and a serine protease gene (Yamaguchi *et al.* 2002). A third C3-like exoenzyme from *S. aureus* has also been discovered from sequencing of the exfoliative toxin B (ETB) plasmid and named EDIN C or C3stau3 (Yamaguchi *et al.* 2001). C3stau3 shares ~65% sequence identity with EDIN A and C3stau2 (see Figure 2.3, Chapter 2) and has been shown to ribosylate RhoA more efficiently than EDIN *in vitro* (Yamaguchi *et al.* 2001). So far C3stau3 activity and any possible role in disease has not been widely studied, but strains having the ETB plasmid are more common than those without indicating that it may have a possible role in virulence (Mulligan *et al.* 1993). One local study of the prevalence of these exoenzymes in passive and pathogenic strains of *S. aureus* showed that strains isolated from infection sites had a higher prevalence of EDIN genes than those isolated from nasal carriers (Czech *et al.* 2001). Also, 90% of the EDIN genes found in this study were C3stau2 (EDIN B), probably due to its chromosomal location. Although it was reported that C3stau3 (EDIN C) is more common in Japan. This data has not been correlated with type of infection however.

C3stau2 was highlighted for investigation due to the finding that it not only ADP-ribosylates RhoA-C, but also RhoE/Rnd3 on the equivalent residue, Asn 44 (Wilde *et al.* 2001). RhoE/Rnd3 is 50% identical to RhoA but lacks GTPase activity and is therefore constitutively active (Foster *et al.* 1996) (Nobes *et al.* 1998). RhoE/Rnd3 modification was observed *in vitro* and *in vivo* but was noted to be much slower possibly due to its 'GTP bound' conformation, as RhoA-GTP is a poorer substrate for C3bot1 than RhoA-GDP (Wilde *et al.* 2000). Mutational analysis of C3stau2 has assessed the importance of key residues involved in NAD-binding and ribosyltransferase activity in C3bot1 (Wilde *et al.* 2002c). Mutants were tested for NAD glycohydrolase activity and ADP-ribosyltransferase activity to distinguish between these two roles. R48K, R85K and E180L all showed decreased glycohydrolase activity as well as decreased transferase activity showing that these residues are probably involved in NAD binding. This would be expected for the catalytic residue, Glu 180, and the equivalent residues in C3bot1 to Arg 48 and Arg 85

appear to be involved in NAD binding in the crystal structure (Menetrey *et al.* 2002). Mutation of the exposed Tyr 175 in the ARTT loop decreased transferase activity but actually increased glycohydrolase activity suggesting that it plays a part in RhoA recognition, although the increase in glycohydrolase activity is unexplained. Ser 137 is the first serine in the conserved 'STS' motif, which in C3stau2 is actually STQ. This residue was originally speculated to be involved in stabilising the catalytic glutamate in C3bot1 but mutation studies have since shown it to have no role in transferase activity (Menetrey *et al.* 2002). Mutation of this residue showed no decrease in transferase activity suggesting that it has no role to play in catalysis in C3stau2 either. Finally, Gln 212 in C3bot1, which was originally speculated to be involved in RhoA recognition from the C3bot1 structure but has now been shown not to be involved in NAD hydrolysis or RhoA binding (Menetrey *et al.* 2002) is equivalent to Gln 178 in C3stau2. Mutation of this residue causes no decrease in NAD hydrolysis or RhoA recognition, as in C3bot1, but in both proteins its mutation knocks out transferase activity suggesting it is involved in stabilising the C3-Rho-NAD complex.

In order to further characterise this protein and to address the question of substrate specificity for the C3 exoenzymes we have crystallised and solved the structure of the native and the NAD bound forms of C3stau2 to 1.68Å and 2.02Å respectively. More recent results show that EDIN A also ribosylates RhoE/Rnd3 (Wilde *et al.* 2003) suggesting that this structure characterises the new subfamily of C3-like exoenzymes from *S. aureus*.

4.2 Methods

4.2.1 Cloning and expression

The cloning was carried out by Dr. J. Mark Sutton at the Health Protection Agency, Porton Down. A synthetic gene encoding C3stau2 was synthesised using codon bias optimised for expression in *E. coli*. A series of overlapping oligonucleotides was synthesised (Sigma-Genosys; UK) as Table 4.1. The numbering indicates the order of the oligonucleotides with the forward primers numbered 5'-3' along the sense strand and the reverse primers numbered 5'-3' along the anti-sense strand. In addition to the structural

Table 4.1: Oligos synthesised to create synthetic C3stau2 gene.

Oligo name	Sequence
C3s1. for	GCTGAAACCAAAAACTTCACCGACCTGGTTGAAGCTACCAAATGG GGTAAC TCTCTGATCAAATCTGCTAAATACTCTTC
C3s2. for	TAAAGACAAAATGGCTATCTACAACTACACCAAAAACTCTTCTCCG ATCAACACCCCGCTGCGTTCTGCTAACGGTGACGTTAACAAACTG
C3s3. for	TCTGAAAACATCCAGGAACAGGTTCGTCAGCTGGACTCTACCATCT CTAAATCTGTTACCCCGGACTCTGTTTACGTTTACCG
C3s4. for	TCTGCTGAACCTGGACTACCTGTCTTCTATCACCGGTTTCACCCGTG AAGACCTGCACATGCTGCAGCAGACCAACAACGG
C3s5. for	TCAGTACAACGAAGCTCTGGTTTCTAAACTGAACAACCTGATGAAC TCTCGTATCTACCGTGAAAACGGTTACTCTTCTACCC
C3s6. for	AGCTGGTTTCTGGTGCTGCTCTGGCTGGTCGTCCGATCGAACTGAA ACTGGAACTGCCGAAAGGTACCAAAGCTGC
C3s7. for	TTACATCGACTCTAAAGAACTGACCGCTTACCCCGGTCAGCAGGAA GTTCT GCTGCCGCGTGGTACCGAATACGCTG
C3s8. for	TTGGTTCTGTTAAACTGTCTGACAACAAACGTAAAATCATCATCAC CGCTG TTGTTTTCAAGAAG
C3s9. rev	GACAGTTTAACAGAACCAACAGCGTATTCGGTACCACGCGGC
C3s10. rev	GTTCTTTAGAGTCGATGTAAGCAGCTTTGGT ACCTTTCGG
C3s11. rev	GAGCAGCACCAGAAACCAGCTGGGTAGAAGAGTAACCGTTTTCAC
C3s12. rev	CCAGAGCTTCGTTGTACTGACCGTTGTTGG TCTGCTGCAGC
C3s13. rev	CAGGTAGTCCAGGTTCAGCAGACGGTAAACGT AAACAGAGTCC
C3s14. rev	GTTCCTGGATGTTTTCAGACAGTTTGTTAACGTCACCGTTAGC
C3s15. rev	GTAGATAGCCATTTTGTCTTTAGAAGAGTATTTAGCAGATTTGATC AG
C3sBam Xa. for	GGGATCCATCGAAGGTCGTGCTGAAACCAAAAACTTCACCG
C3sBam rev	CGGATCCTTATCACTTCTTGAAAACAACAGCGG

oligonucleotides, flanking primers were designed to introduce a 5'and a 3' *Bam*HI site and sequence encoding a Factor Xa cleavage site immediately prior to the initial alanine of the mature C3stau2 peptide sequence. The oligonucleotides at a concentration of 25nmol/µl for the structural primers and 500nmol/µl for the flanking primers were assembled in a single PCR reaction using VENT polymerase (New England Biolabs) in a buffer containing 10mM KCl, 20mM Tris-HCl (pH 8.8 at 25°C), 10mM $(NH_4)_2SO_4$, 4.5mM $MgSO_4$, 0.1% Triton X-100, 1mM dNTPs. PCR amplification of 30 cycles was carried out using the parameters 94°C 1 min, 56°C 1 min, 72°C 1 min followed by 72°C 10mins. The PCR product was analysed on a 1.2% agarose gel in TAE electrophoresis buffer and a principal band at approximately 750 base pairs was excised from the gel. The band was extracted from the gel using Qiaex II gel extraction resin (Qiagen) and subcloned into the sequencing vector PCR4.0 TOPO (Invitrogen) according to manufacturer's instructions. The clone was sequenced and showed two point mutations in the regions of primer C3sBamXa.for and C3s10.rev. These were corrected by amplification with these primers, followed by extension to give the full-length clone and re-amplification with the flanking primers. Cloning and sequence verification were carried out as described above.

For expression of the C3stau2 fragment a modified *malE* fusion vector was generated to ensure that the gene could not be disseminated to other bacteria. The *Apa*I-*Hind*III fragment from the expression vector pMALc2x (New England Biolabs) was isolated and subcloned into the vector pBC SK+ (Stratagene) to generate the vector pBCmalE. The construction of the vector was confirmed by restriction digest and sequencing. A *Bam*HI fragment containing the C3stau2 coding sequence was subcloned into pBCmalE digested *Bam*HI and the clone verified. For expression the clone TB1 pBCmalE C3stau2 was grown overnight at 30°C in Terrific Broth supplemented with 35 µg/ml chloramphenicol and 0.5% (w/v) glucose. The overnight culture was diluted 1:10 in fresh media and grown for 4 hours (OD_{600} ~3.9). The culture was induced with IPTG at a final concentration of 500 µM and grown for a further 2 hours 30 minutes at 25°C before harvest. The cells were collected by centrifugation and resuspended in 20mM MES pH5.8. The cells were lysed by sonication (5 burst of 40s at 26microns, resting on ice between bursts –Soniprep 150) in batches corresponding to half a litre of culture and the cell lysate was collected by centrifugation.

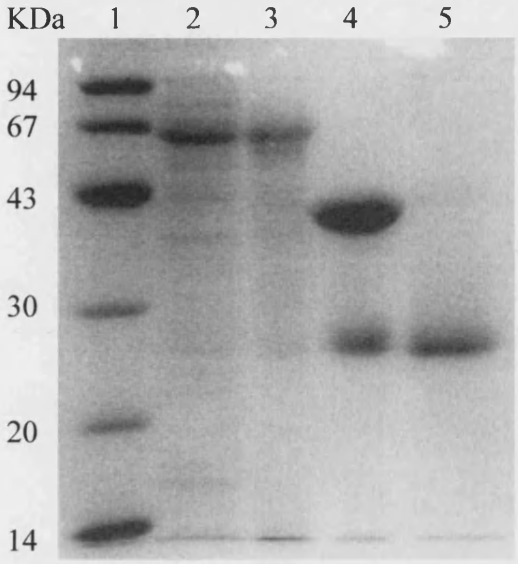


Figure 4.1: 12% SDS-PAGE gel showing purification stages of C3stau2. Lane 1 is the low molecular weight marker (Amersham), lane 2, the soluble extract from the lysed cells, lane 3, the first peak purified from the SP sepharose column at pH 5.8, lane 4, the cleaved fusion protein and lane 5, the C3stau2 after purification with the SP sepharose column at pH 7.3. The molecular weights for C3stau2-MBP, MBP and C3stau2 ~66.0KDa, 42.4KDa and 23.6KDa respectively.

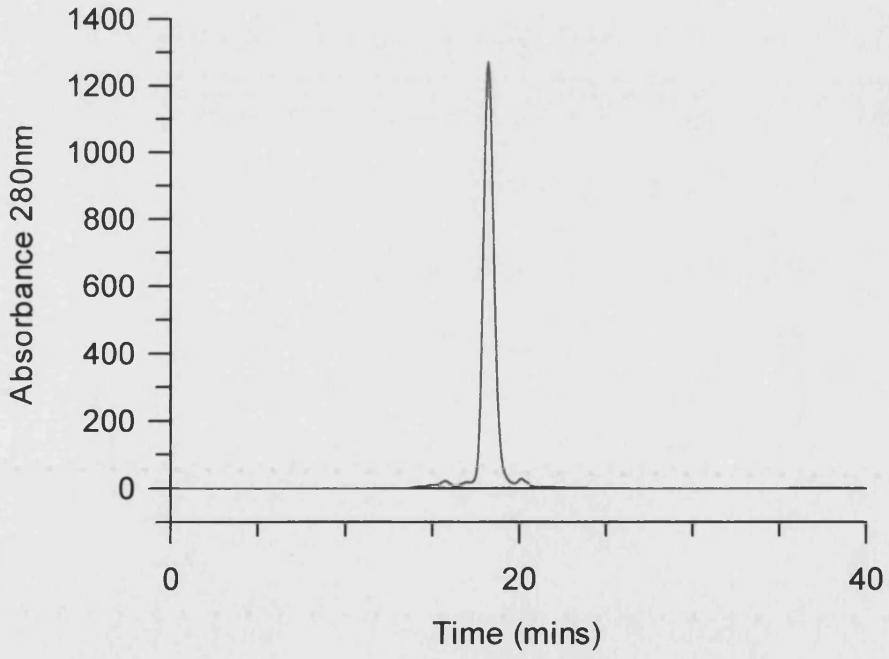


Figure 4.2: Final purification of C3stau2. Chromatograph from the final purification by gel filtration using a Superdex 200 HR 10/30 (Amersham) column run at 0.75ml/min. Fractions corresponding to the main peak were collected and concentrated.

4.2.2 Protein purification

The C3stau2-MBP fusion protein was initially purified by cation exchange chromatography. The protein bound to a SP-sepharose column optimally equilibrated with 20mM MES pH 5.8, and was eluted on an ascending NaCl gradient as a single peak. The fractions corresponding to the peak were pooled and 150U Factor Xa were added per half litre of cells to cleave the MBP as the cleavage reaction was very inefficient. The initial yield of C3stau2-MBP was ~50mg, corresponding to ~17mg C3stau2. During cleavage, the protein was dialysed for 24hrs at room temperature against a buffer containing 20mM Hepes pH 7.3, 50mM NaCl and 5mM $CaCl_2$ to improve the cleavage efficiency and to raise the pH for further purification. The cleaved C3stau2 bound to the SP-Sepharose column equilibrated in 20mM Hepes pH 7.3 and 50mM NaCl and was eluted again, mostly pure, on an ascending NaCl gradient (Figure 4.1). Final purification and de-salting was achieved by gel filtration. The protein was concentrated in an 10kDa cut off centrifugal concentrator (Amicon) and applied to a Superdex 200 gel filtration column (Amersham) equilibrated in 20mM Hepes pH7.3, 50mM NaCl (Figure 4.2). The purified protein was finally concentrated to 35mg/ml and stored in 20mM Hepes pH 7.3, 50mM NaCl at -70°C. The final yield of C3stau2 from half a litre of cells was ~7mg.

4.2.3 Crystallisation

The solubility of C3stau2 was initially characterised using commercial screens by the vapour diffusion hanging drop technique. 1µl of the well solution was mixed with 1µl protein to conserve material. Initial experiments carried out at 22°C and 16°C at varying protein concentrations suggested that a protein concentration of 8-10mg/ml was suitable

Figure 4.3: C3stau2 crystals. Orthorhombic crystals for C3stau2 grew in two days (after seeding) at 22°C using the hanging drop technique at a protein concentration of 8.75mg/ml. 1µl C3stau in 20mM Hepes pH 7.3, 50mM NaCl was mixed with 1µl well solution (30% PEG 8000, 0.1M Sodium Cacodylate pH6.5).

for further screening at 22°C. As commercial screens did not produce preliminary crystals, grid screens were designed to test a range of precipitants to find suitable conditions for further screens. C3stau2 concentrations of 8.75-10.25mg/ml produced large single orthorhombic crystals in 29-30% PEG 8000, 0.1M Sodium cacodylate pH 6.4-6.6 at 22°C in three weeks. Not all drops produced large crystals so some drops were streak seeded. Seeding produced larger numbers of good sized crystals which grew within 3 days and were suitable for cryo-diffraction (Figure 4.3). Co-crystallisation experiments with either 10mM nicotinamide or ADP-ribose were also performed. Additionally, crystal soaks were performed with NAD, nicotinamide and ADP-ribose. Neither nicotinamide or ADP-ribose were observed bound to these crystals, which may be due to the pH of the crystallisation. The optimal pH for the forward enzyme reaction, hydrolysing NAD, as assessed for C3bot1, is pH7.5, whereas it is pH 5.5 for the reverse reaction (Habermann *et al.* 1991). This suggests that nicotinamide and ADP-ribose are probably more likely to bind at this pH. Unfortunately, soaking these crystals at pH 5.5 caused them to dissolve.

4.2.4 Data collection and processing for native C3stau2

A room temperature data set from a large crystal was collected first using a ADSC-CCD detector on station PX 14.1 at the Synchrotron Radiation Source at Daresbury, UK. The data was processed and scaled in HKL2000 (Otwinowski and Minor 1997) to a resolution of 2.62Å and the symmetry and systematic absences were consistent with the space group $P2_12_12_1$. The Matthews coefficient (Matthews 1968) of 2, indicated that there was one C3stau2 monomer per asymmetric unit and the crystals contained 40% solvent. Data reduction was carried out using the program TRUNCATE (CCP4 1994). Data processing statistics for all data sets collected are shown in Table 4.1. From this data set, a molecular replacement solution was found with MOLREP (Vagin and Teplyakov 1997) using residues 47-246 of the crystal structure of the native C3bot1 (Han *et al.* 2001). MOLREP was run several times using different resolution limits and the most common solution ranked first had an R-factor of 56% and a correlation coefficient of 31.4. However, initial refinement, carried out using CNS version 1.0 (Brünger *et al.* 1998) showed it to be correct.

The large structural differences between C3stau2 and C3bot1 made it hard to discern the position of the extra loops in the density so a further data set under cryo conditions was collected. Native crystals, soaked in a cryoprotectant containing 0.1M Sodium Cacodylate, pH6.5, 30% PEG 8000 and 20% glycerol and cryo-cooled to 100K, diffracted to 1.79Å on a ADSC-CCD detector on station PX 14.2 at the Synchrotron Radiation Source at Daresbury, UK. Later, crystals co-crystallised with 10mM nicotinamide were cryocooled and diffracted to a resolution of 1.68Å on a ADSC-CCD

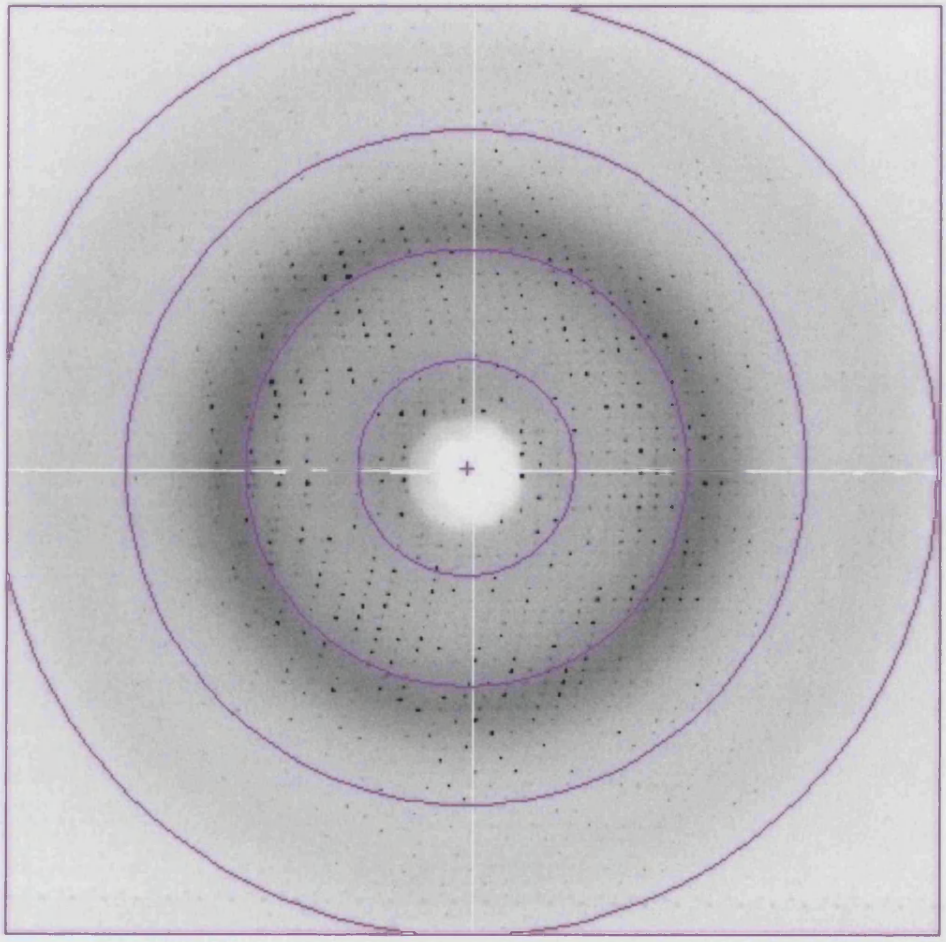


Figure 4.4: **Diffraction image for C3stau2 from the data set scaled to 1.68Å**. The rings mark the resolution limits 6.71Å, 3.31Å, 2.21Å and 1.65Å proceeding outwards from the centre. The reflections are clearly defined and lunes are visible on the left hand side. The data were collected on station PX 9.6 at the SRS, Daresbury, U.K. The wavelength was 0.87Å and the exposure time 12 seconds.

detector on station PX 9.6 at the Synchrotron Radiation Source at Daresbury, UK. These crystals were shown to have no nicotinamide bound and were used as the highest resolution native data set. Using this data set all the new residues could be built in. For all of the data sets collected, the reflections were well defined and showed a clear, non-mosaic pattern (Figure 4.4).

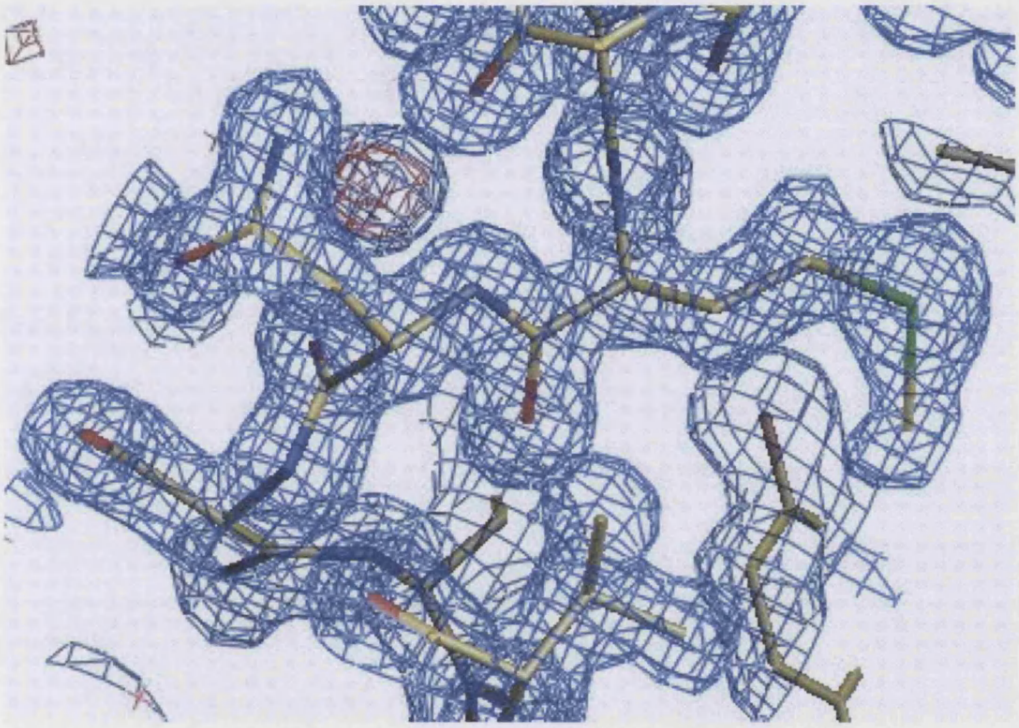


Figure 4.5: **Final map and model for the C-terminal end of** α**7**. The C3stau2 model is shown as a stick representation, the $2F_o$-F_c map at 1σ level is represented as a 3D blue mesh and the F_o-F_c map at 3σ level as a red mesh. This picture was created in O (Jones *et al.* 1991).

4.2.5 Refinement and Model Building for native C3stau2

The refinement was performed in three stages, refinement of the 2.6Å data set, followed by increasing the resolution to 1.79Å, allowing the building of the extra loops of C3stau2, not in C3bot1, before continuing the refinement with the maximum resolution data set at 1.68Å. The initial refinement of the 2.6Å, 1.79Å and 1.68Å data sets was carried out using the program CNS (Brünger *et al.* 1998). In the later stages of the refinement of the 1.68Å data set some water molecules were added. Insertion of further water molecules and final refinement were performed in SHELX (Sheldrick and Gould 1995) to aid the

addition of alternate conformations. The water positions were checked to have heights above 3σ in the F_o-F_c map and to have a temperature factor less than 40Å^2. The behaviour of the R_{free} value (872 reflections) was monitored throughout the refinement. Refinement statistics are shown in the crystallographic statistics in Table 4.3. The program O (Jones *et al.* 1991) was used to rebuild the model and visualise the maps. The final model contains 1660 protein atoms and 284 water molecules has a crystallographic R_{cryst} of 17.0% and an R_{free} value of 23.7%. At 1.68 Å, the density for the loop regions of C3stau2, not seen in C3bot1, are well defined (Figure 4.5). The N and C termini were clearly visible in the structure but residues 197-199 appeared slightly disordered at this resolution. The extremities of side chains of 85, 94, 116 and 200 have been modelled with an occupancy of zero due to insufficient density. The structure was checked with PROCHECK (CCP4 1994) and the Ramachandran plot showed that 90% of the residues were in the most favourable regions of the plot and only one, which was disordered, was in a disallowed region (Figure 4.6).

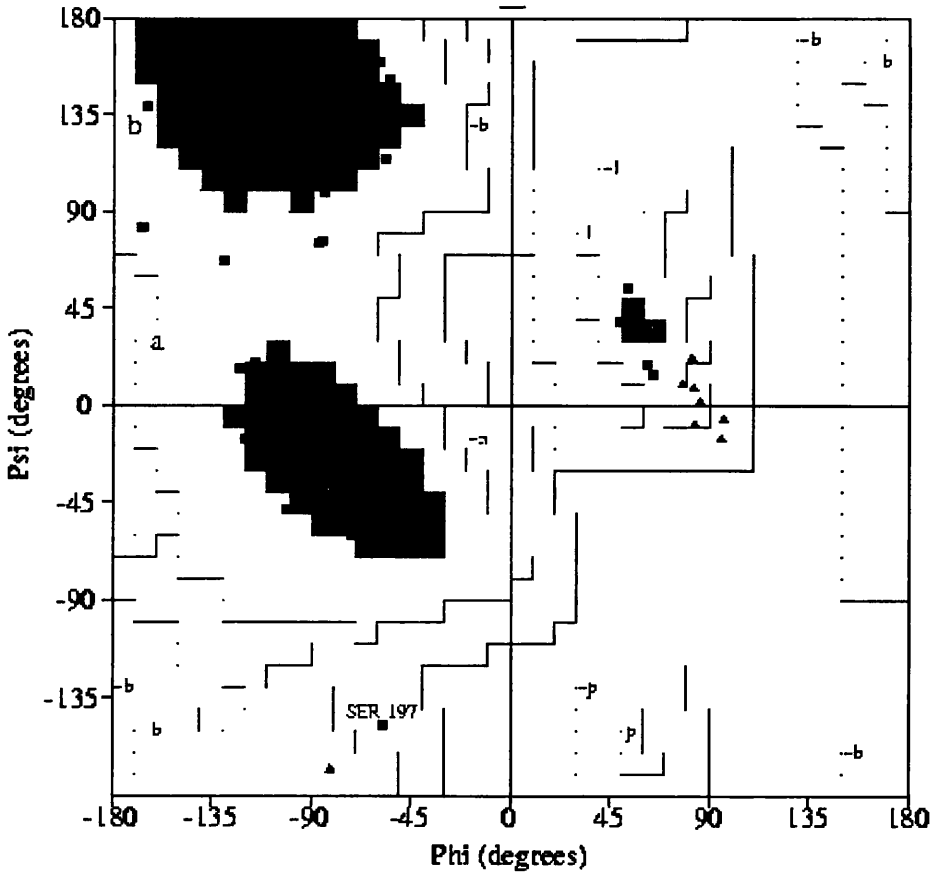


Figure 4.6: Ramachandran plot for C3stau2.

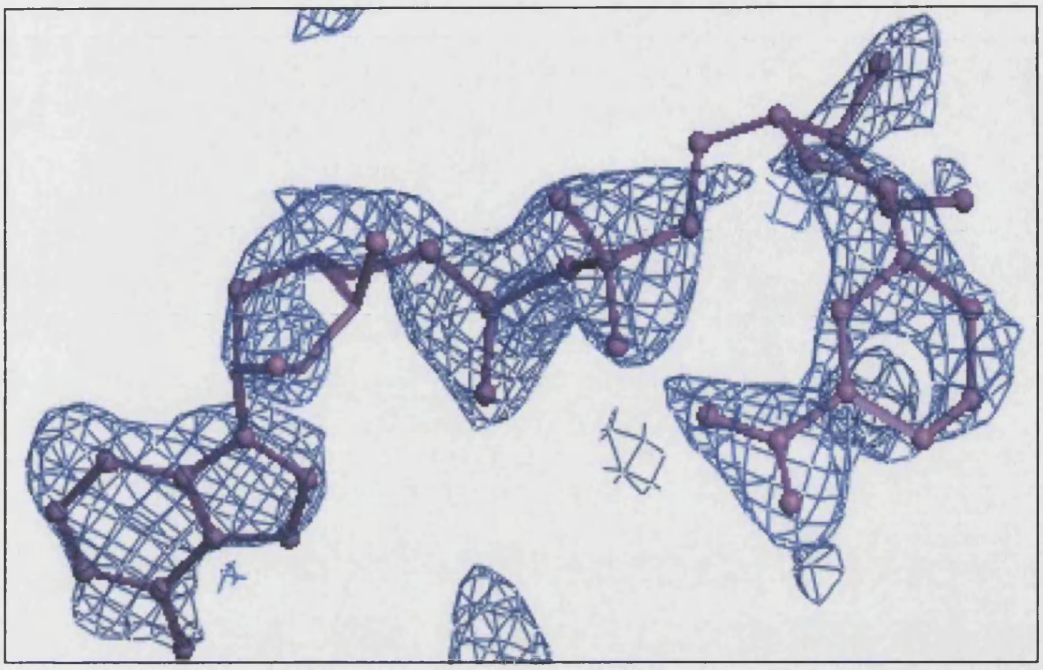


Figure 4.7: Initial positioning of NAD in the F_o-F_c density at 2σ level. Figure was created using O (Jones *et al.* 1991). For NAD atoms see Figure 2.1, Chapter 2.

4.2.6 Data collection and processing for C3stau2-NAD complex

Data sets were collected for crystals soaked in nicotinamide and ADP-ribose but did not show binding of the substrate. Co-crystallisations with nicotinamide showed none bound, and crystals did not grow with ADP-ribose in the crystallisation solution. Concentrations of substrates tried and data collection statistics are shown in Table 4.2. Soaking native crystals in NAD proved successful however. Native crystals were soaked in 25mM NAD for 1 hour and 11hrs and then in cryoprotectant containing 0.1M Sodium Cacodylate, pH 6.5, 30% PEG 8000 and 20% glycerol before cryo-cooling to 100K. The crystals diffracted to 2.02Å and 2.64Å on stations 14.2 and 14.1 respectively at the Synchrotron Radiation Source, Daresbury, UK. The data were collected using a ADSC-CCD detector. The data were processed and scaled in HKL2000 (Otwinowski and Minor 1997) and data reduction was done with TRUNCATE (CCP4 1994).

4.2.7 Refinement and model building for the C3stau2-NAD complex

As the data were isomorphous with the native data, the phases were taken directly from the native model followed by a round of rigid body refinement. As both sets of data showed the NAD bound, refinement proceeded with the 2.02Å data set. Refinement and water picking were performed using CNS (Brünger *et al.* 1998) and the behaviour of the R_{free} value (925 reflections) was monitored throughout the refinement. The water positions were checked to have heights above 3σ in the F_o-F_c map and to have a temperature factor less than 50Å^2. NAD was included in the refinement as clear difference density for NAD was observed in the F_o-F_c map at 2 σ level (Figure 4.9). Most of the model is well ordered except for the end-most 2-3 atoms of the side chains of residues 30, 56 and 116 which were modelled at zero occupancy. Methionine 105 and asparagine 124 are modelled as double conformations each with 0.5 occupancy. The program O (Jones *et al.* 1991) was used to rebuild the model and visualise the maps. The final model had an R_{cryst} of 21.4% and an R_{free} of 26.9%. The details of data processing and refinement are in Table 4.3. All the following figures were made with Molscript (Kraulis 1991) and rendered with Povray (www.povray.org) unless otherwise stated.

Table 4.1: Data collection statistics

Data set and soaking conditions	Resolu-tion of last shell (Å)	Exposure (sec)	I/sigI (last shell)	Measured reflections	Unique reflections	Completeness (%)	R_{sym}(final shell) (%)[a]
Native	2.71-2.62	6	9.2 (6.6)	32,286	7,272	82.3 (77.0)	7.2 (9.5)
Native	1.85-1.79	10	21.9 (3.3)	337,441	20,739	99.1 (97.9)	7.1 (46.0)
25mM NAD for 1hr	2.09-2.02	10	15.8 (4.1)	327,780	25,780	96.9 (96.6)	8.8 (34.1)
25mM NAD for 11hrs	2.73-2.64	10	13.1 (6.0)	123,240	7,003	99.5 (98.7)	12.6 (27.3)
50mM ADP-ribose for 2hrs	2.76-2.66	10	14.7 (3.6)	155,230	6,959	98.1 (92.6)	11.1 (39.8)
50mM ADP-ribose for 1.5 hrs	2.61-2.52	15	23.9 (9.2)	139,270	8,181	99.4 (99.1)	6.9 (18.6)
50mM nicotinamide for 1.3hrs	2.27-2.19	12	13.8 (6.0)	194,209	20,739	91.6 (87.5)	11.5 (32.4)
Nicotinamide co-crystals	1.74-1.68	12	20.0 (5.3)	409,845	224,78	99.4 (97.5)	8.5 (29.5)

[a] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I(h)$, where $I_i(h)$ and $I(h)$ are the ith and the mean measurements of the intensity of reflection h, respectively.

Table 4.3: Crystallographic statistics

	Native crystals	NAD
Space group	$P2_12_12_1$	$P2_12_12_1$
Maximum resolution (Å)	1.68	2.02
Unit cell dimensions (Å)	a=39.62 b=64.53 c=74.98	a=41.90 b=65.08 c=75.07
Measured reflections	409,845	327,780
Unique reflections	22,478	14,088
Completeness (%) (last shell)[a]	99.4 (97.5)	96.9 (96.6)
Mean I/σ (last shell)[a]	20.0 (5.3)	15.8 (4.1)
R_{sym} (%)[b] (last shell)[a]	8.5 (29.5)	8.8 (34.1)
R_{cryst} (%)[c]	17.0	21.4
R_{free} (%)[d]	23.7	27.3
Number of protein atoms	1660	1660
Number of solvent atoms	284	154
Number of ligand atoms		41
R.m.s. deviation in bond lengths (Å)	0.004	0.006
R.m.s. deviation in bond angles (°)	1.3	1.3
Average B factor for protein atoms (Å^2)	15.3	24.5
Average B factor for solvent molecules (Å^2)	19.0	27.3
Average B factor for ligand atoms (Å^2)		31.1

[a] Last shell is 1.74-1.68 and 2.09-2.02 for the native and NAD complexes respectively.

[b] $R_{sym} = \Sigma_h \Sigma_i |I(h) - I_i(h)| / \Sigma_h \Sigma_i I(h)$, where $I_i(h)$ and $I(h)$ are the ith and the mean measurements of the intensity of reflection h, respectively.

[c] $R_{cryst} = \Sigma_h |F_o - F_c| / \Sigma_h F_o$, where F_o and F_c are the observed and calculated structure factor amplitudes of reflection h, respectively.

[d] R_{free} is equal to R_{sym} for a randomly-selected 850-950 reflections not used in the refinement (Brünger 1992).

4.3 Results and Discussion

4.3.1 Overall Topology of C3stau2

The overall structure of C3stau2 is similar to that of C3bot1, i.e. a mixed α/β fold with a β-sandwich core (Figure 4.8A) (Han *et al.* 2001; Menetrey *et al.* 2002). Structural alignment of C3stau2 with C3bot1 (Figure 4.8B) shows that the main differences are restricted to the region of helices α5 and α6 (residues 92-126) in C3stau2 which corresponds to the loop between β1 and α5 in C3bot1 (residues 135-162) (Figure 3.5). This region in C3stau2 includes seven extra residues that contribute to two extra helices comprising residues 99-108 (α5) and 115-125 (α6). This extended variable region is linked

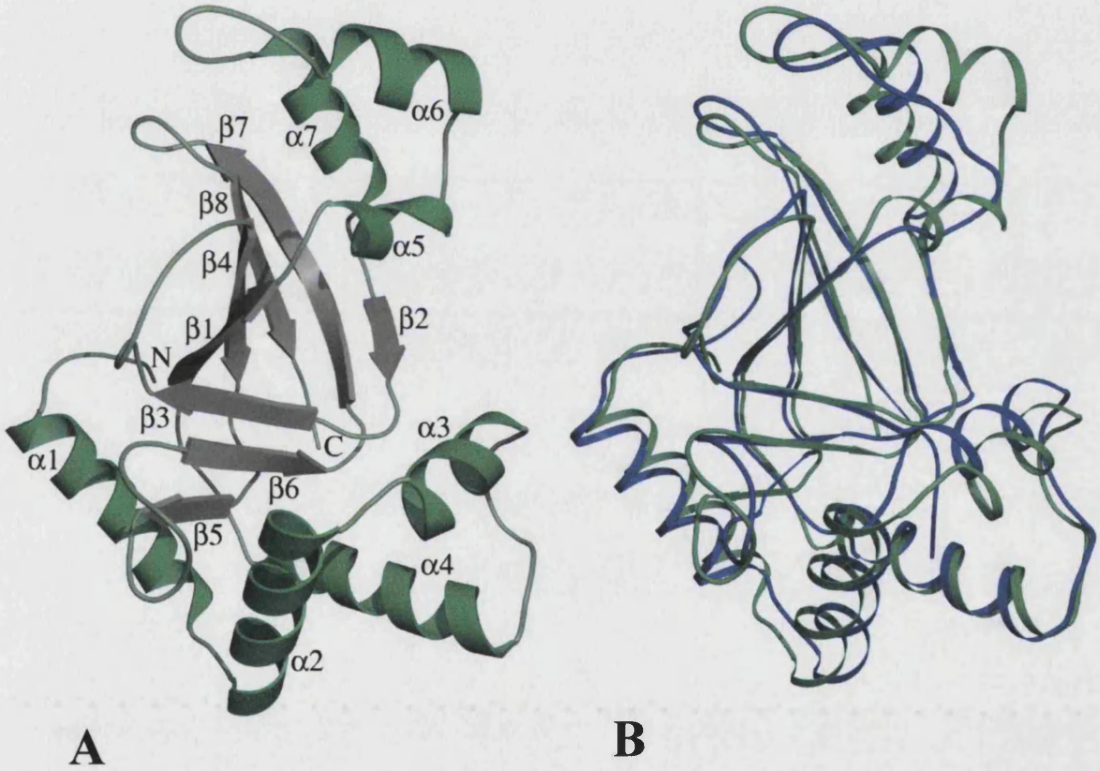


Figure 4.8: A: Overall structure of C3stau2 (PDB=1OJQ). Schematic representation in which helices, strands and main chain termini are labelled. **B: Superposition of C3stau2 (PDB=1OJQ) and C3bot1 (PDB=1G24).** C3stau2 (Evans *et al.* 2003) is shown in green and C3bot1 (Han *et al.* 2001) is shown in blue. Alignment of the two structures was carried out using LSQ routines within O (Jones *et al.* 1991).

by main chain hydrogen bonding to the loop between strands β7 and β8 (residues 197-202) which also deviates from its counterpart in C3bot1 (r.m.s. deviation of Cα atoms is 1.6Å). Here, the main chain oxygens of residues Asn 199 and Lys 200 are bonded to the main chain nitrogen of Tyr 114 and the side-chain nitrogen of Gln 133 respectively.

NAD was added to the second model after initial refinement against the native C3stau2 structure showed clear difference density at 2σ level. The complexed form has the same topology as the native form of C3stau2 with a Cα r.m.s. deviation of 0.83 Å.

4.3.2 The ARTT loop

The bacterial ADP-ribosylating toxins all share a conserved glutamate that is essential for ADP-ribosylation. In C3stau2, this residue (Glu 180) is located between strands β5 and β6 at the end of the second turn of what has been termed for C3bot1, the 'ADP-ribosylating turn turn', or ARTT, motif. The ARTT loop been shown to be essential for binding the NAD and positioning the C3-NAD-Rho complex for ADP-ribosyl transfer by mutational analysis (Menetrey *et al.* 2002; Wilde *et al.* 2002b; Tsuge *et al.* 2003) and was shown to change orientation in C3bot1 on NAD binding (Menetrey *et al.* 2002). In particular, three residues of the ARTT loop have been shown to be essential for C3stau2 ADP-ribosylation. As well as the catalytic glutamate on turn two, there is a conserved glutamine (or glutamate) thought to be essential for positioning the ternary C3-NAD-Rho complex on turn two and a conserved tyrosine (or phenylalanine) on turn one necessary for Rho binding (Wilde *et al.* 2002b).

Despite the conserved role of the ARTT loop in ADP-ribosylation, the loop is orientated differently in C3stau2 and C3bot1. In the native structure of C3stau2, the backbone of the first turn of the ARTT loop is shifted by up to 5.1Å compared to the backbone of the same turn in C3bot1 (Figure 4.9), although this does not affect the conserved position of Glu 180. The three C3-like exoenzymes from *S. aureus* are unusual amongst the C3-like and binary toxins discovered so far in that the loops immediately prior to their ARTT loops contain an extra two residues. These two residues appear to be responsible for positioning the ARTT loop in a conformation identical, not to that of the native C3bot1 structure, but to that of the C3bot1-NAD complex, the 'NAD bound'

conformation. The loop is well ordered with B-factors (mean =10.2Å^2 over 8 Cα atoms) below average (15.3 Å^2) for the structure and appears optimally placed for ribosylation.

It is not surprising then, that upon binding of NAD, the ARTT loop undergoes just a slight conformational change (Figure 4.9). The r.m.s. deviation for the ARTT loop between the native and the NAD-complexed structures over 8 Cα atoms is 0.1 Å. As expected, Glu 180 can be seen hydrogen bonding to the nicotinamide ribose nitrogen, NO2′ (Figure 4.10), an interaction for which it does not need to move. In C3stau2, the position of Glu 180 is not stabilised by hydrogen bonding to the PN loop as seen in C3bot1. However, Glu 180 does interact with Ser 138, the first serine in the STX motif (also known as the STS motif), as seen in C3bot1. The YX STX motif has been identified as a conserved motif across the ADP-ribosyltransferase toxins and in the case of the C3 exoenzymes from *S. aureus* this motif has the sequence YS STQ. This motif has generally been considered to help stabilise the NAD binding pocket and mutational analysis of this motif in diphtheria toxin and cholera toxin has helped distinguish differences in the NAD

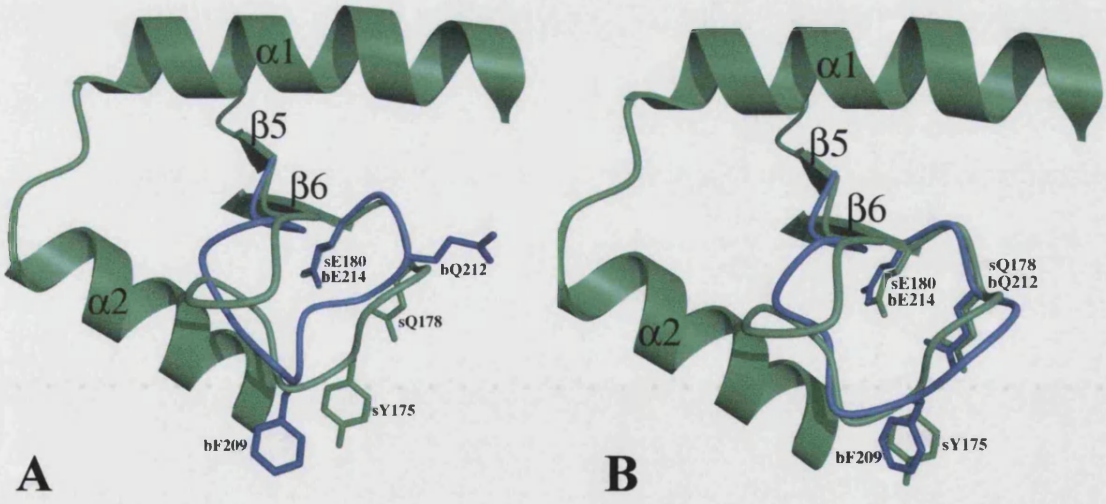


Figure 4.9: A: Superposition of the ARTT loops of C3bot1 (PDB=1G24) and C3stau2 (PDB=1OJQ) (Han *et al.* 2001; Evans *et al.* 2003). C3stau2 is in green and C3bot1 in blue. Residues from C3stau2 are prefixed 's' and those from C3bot1 'b'. **B: Superposition of the ARTT loops of the C3bot1-NAD complex (PDB=1GZF) and C3stau2 (PDB=1OJQ)** (Menetrey *et al.* 2002; Evans *et al.* 2003). The C3stau2 ARTT loop rests in the 'NAD-bound' conformation.

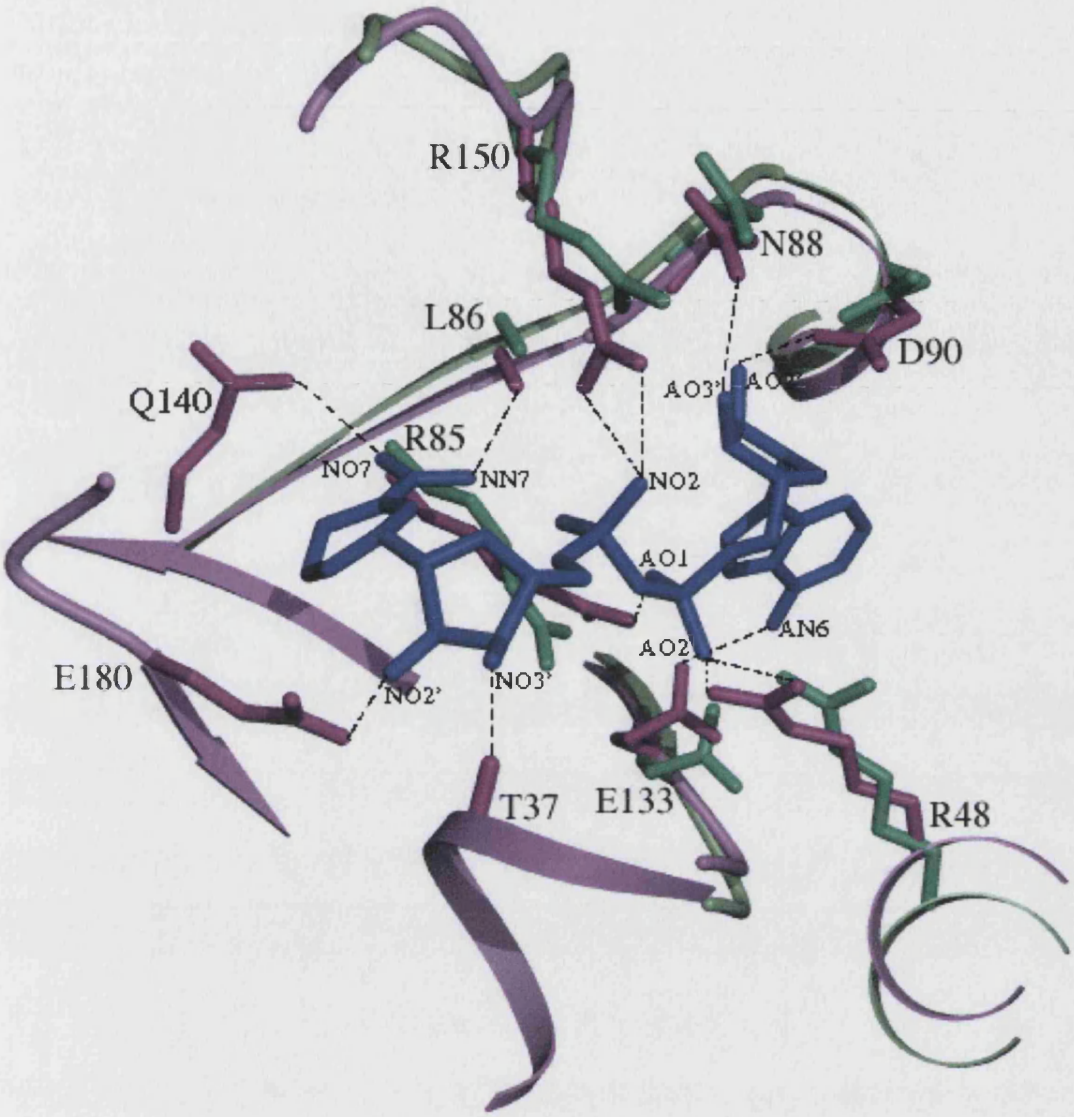


Figure 4.10: Movement of residues involved in binding NAD. The native C3stau2 is shown in green and the C3stau2-NAD complex is shown in pink with the NAD in blue. The large domain movements involved in binding NAD have moved residues R85, L86, R150, N88, D90, R48 and E133 within binding distance of the NAD. E180, T37 and Q140 also bind NAD. For clarity, the interactions between the main chain nitrogens of L86 and N88 and the NO7 and AO2′ atoms of the NAD respectively are not shown.

hydrolysis mechanism between the two types of A-B toxins (Dolan *et al.* 2000). Also, mutation of the first serine of the STX motif has shown it to be essential for NAD hydrolysis and ADP-ribosylation in the *C. difficile* Iota (binary) toxin (Nagahama *et al.* 2000) and to be likely to help stabilise the transition state with the catalytic glutamate (Sakurai *et al.* 2003). However, mutation of this serine in C3bot1, Ser 174 (Menetrey *et al.* 2002), and C3stau2, Ser 138 (Wilde *et al.* 2002c), has very little effect on NAD binding or catalysis. Gln 140, however, the last C3stau2 residue in this motif, is positioned to hydrogen bond to the NO7 atom of NAD. These data may suggest that the role of this motif may vary between ADP-ribosyltransferases depending upon the contribution of the other residues in the NAD binding pocket to transition state stabilisation.

The 'NAD bound' conformation of the native ARTT loop also has implications for Gln 178. This residue, and its homologues in C3bot1 and the *C. difficile* Iota toxin, have been shown by mutational analysis to have a role in ADP-ribosylation, but not in isolated NAD or Rho binding (Menetrey *et al.* 2002; Wilde *et al.* 2002c; Sakurai *et al.* 2003; Tsuge *et al.* 2003). This key residue, which is thought to stabilise the C3-NAD-Rho ternary complex, is pointing towards the NAD binding cleft in a position identical to that of its equivalent residue in the C3bot1-NAD complex, Gln 212 (Figure 4.9). Gln 178 does not move on NAD binding or hydrogen bond to the NAD confirming its non-essential role in NAD binding.

The other important residue at the end of turn 1 of the ARTT loop is Tyr 175. This tyrosine residue, and its homologue in C3bot1, Phe 209, are the only residues so far implicated in RhoA binding. The solvent exposure of C3bot1 Phe 209 and its location near to the active site residues suggested a role in binding to a hydrophobic patch on RhoA (Han *et al.* 2001) and the necessity of C3stau2 Tyr 175 for RhoA binding has been shown by mutational analysis (Wilde *et al.* 2002c). Once again, the position of this residue in native C3stau2 resembles that of its counterpart in NAD-complexed C3bot1 and does not move on NAD binding.

These results lend support to one obvious conclusion: the resting position of the ARTT loop in different C3-like exoenzymes can resemble to varying extents the conformation observed in the NAD-containing complex. The loop may, of course, rearrange further on ternary complex formation or during catalysis.

4.3.3 The PN binding loop

The structure of C3bot1 in complex with NAD identified a further loop between strands β3 and β4, comprising residues 141-150 in C3stau2, termed the phosphate-nicotinamide (PN) loop (Menetrey *et al.* 2002). This loop is so-called as it appears to be involved in binding the nicotinamide and the nicotinamide-phosphate moieties of NAD through provision of a hydrophobic pocket. This loop was noted to alter its shape dramatically on the binding of NAD to C3bot1 (Menetrey *et al.* 2002). Interestingly, in C3stau2, this short loop also appears in a conformation different to that found in native C3bot1. In C3stau2, the loop appears more extended (Figure 4.11) with a r.m.s. deviation of 2.5 Å over ten Cα atoms when compared to C3bot1. Although this loop contains two more alanines than in C3bot1, making a total of three alanines and two glycines, the mean B-factor for this loop (16.1 $Å^2$) is not significantly higher than the mean for the whole molecule (15.3 $Å^2$). The loop also makes three crystal contacts with symmetry-related molecules which may contribute to its stability in the crystal structure. It is therefore possible that this loop is more flexible in solution prior to NAD binding.

Figure 4.11 shows that in contrast to the ARTT loop, the PN binding loop moves towards NAD on binding with a change in position corresponding to an r.m.s. deviation of 1.7 Å over 10 Cα atoms. C3stau2 does not have a residue equivalent to Phe 183 in C3bot1 which stacks with the adenine ring, but provides a hydrophobic pocket of alanines, glycines and Leu 148. The result of this conformational change is that Arg 150 is moved by approximately 1 Å closer to the NAD, allowing binding with the NO2 atom of NAD (Figure 4.11).

4.3.4 C3stau2 moves to enclose the NAD

The structure of C3stau2 shows significant conformational change on binding NAD. The structure of C3bot1 in complex with NAD, when aligned along helices 2-4, shows a movement of a large domain comprising β2, α5, β7 and β8 towards the NAD binding cleft, closing the protein structure around the NAD. Figure 4.11 shows that C3stau2 also undergoes a similar clasping movement involving the variable domain (helices α5 and α6, along with strands β7 and β8) and helix α3 to enclose its substrate. The variable domain can be seen to rotate up to 4° about the catalytic residue, Glu 180,

towards the binding site. As with C3bot1, this change distorts the C-terminal end of strand β1 towards the NAD, in this case, bringing Arg 85 and Leu 86 into range for NAD binding. Also, helices α3 and α4 on the other side of the NAD binding cleft can be seen rotating a similar amount in the opposite direction to close the gap further.

Figure 4.11: Superposition of native C3stau2 (PDB=1OJQ) and the C3stau2-NAD (PDB=1OJZ) (Evans *et al.* 2003) complex in two orientations. Native C3stau is in grey and the domains that have moved to bind NAD in the complex are shown in pink. The PN binding loop is shown in green, NAD is in blue. The first orientation (left) shows the change in the conformation of the PN binding loop most clearly. The second orientation (right) best shows the clasping of the domains to enclose the NAD in the binding site. The insert shows how the change in orientation of the PN binding loop allows R150 to bind to the NAD.

Figure 4.10 shows that these large structural changes bring the α carbons of seven of the ten NAD binding residues between 0.82 Å and 1.51 Å closer to the NAD. This, in turn, allows the critical side chains to be brought within bonding distance of the NAD. Most notable are residues Arg 85 and Arg 48 which have been shown by mutational analysis to be essential for NAD binding in C3stau2 (Wilde *et al.* 2002c). A change in the

side chain conformation and a main chain Cα shift of 0.82 Å in Arg 85 on strand β1 moves its NH2 atom 2.95 Å, allowing interaction with the adenosine phosphate AO1 atom of NAD. For Arg 48 on helix α3, a change in side chain conformation and a Cα shift of 1.51 Å moves its NH1 atom 2.16 Å, allowing bonding with the NAD AO2 atom.

4.3.5 Analysis of the common NAD-binding site of the C3-like and binary toxins

The structures of the C3-like and binary toxins determined so far reveal similar, but not identical NAD binding sites. Table 4.4 shows the residues of C3stau2, C3bot1 (Menetrey *et al.* 2002), and the binary toxins, Iota toxin from *C. perfringens* (Sakurai *et al.* 2003; Tsuge *et al.* 2003) and VIP2 from *B. cereus* (Han *et al.* 1999), that are involved in binding NAD/NADH. Superposition of the binding sites from all four structures shows that the nicotinamide portion of the NAD is in a similar orientation with the central phosphates and the adenine differing slightly in their positions. The NAD binding site is characterised not only by the ARTT and PN loops but also by two arginines conserved in the C3-like and binary toxins, and a third conserved in the C3-like toxins.

The most conserved feature of the binding site is the positioning of the nicotinamide moiety by the main chain nitrogen and oxygen atoms of the residue following the conserved arginine equivalent to Arg 85 in C3stau2 (Position 2, Figure 2.3, Chapter 2). As the identity of this residue varies from glycine in C3bot1 to tryptophan in VIP2, it is likely that any residue that can adopt the appropriate backbone conformation will support this interaction. Apart from this residue and the catalytic glutamate, the only other conserved interaction is some stabilisation/orientation of the adenine ring which is positioned by hydrogen bonds in C3stau2, C3bot1 and the Iota toxin and by hydrophobic packing in C3bot1 and VIP2.

In C3stau2, three arginines conserved amongst the C3-like toxins (labelled 1-3 in Figure 2.3, Chapter 2) are seen to interact with the central phosphates of the NAD (Figure 4.11). These interactions have been proposed to be involved in increasing affinity for the substrate (Han *et al.* 1999; Sakurai *et al.* 2003). Mutational analyses of these residues

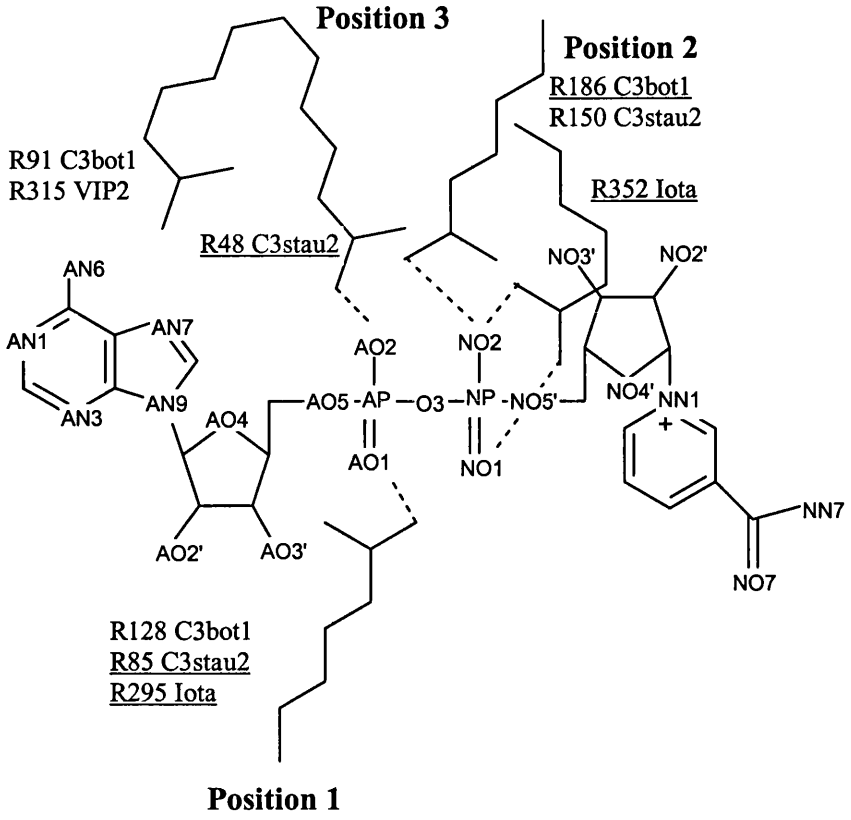


Figure 4.12: Interactions of conserved arginines in C3stau2, C3bot1, Iota toxin and VIP2 with NAD. Residues that interact with NAD from the respective crystal structures (Han *et al.* 1999; Menetrey *et al.* 2002; Evans *et al.* 2003; Tsuge *et al.* 2003) are shown. Underlined residues have been shown to be essential for NAD binding by previous mutational analysis (Menetrey *et al.* 2002; Wilde *et al.* 2002b; Tsuge *et al.* 2003) and potential hydrogen bonds are indicated by dashed lines. Positions 1-3 are the positions of arginine residues, conserved in the C3-like toxins, that interact with NAD in the C3stau2 structure. The arginines at positions 1 and 2 are also conserved in the binary toxins. Their interactions are conserved in the Iota toxin. The nature of the interaction at position 3 varies (VIP2 R315 engages in π–π interactions with the adenine ring whereas C3bot1 R91 Cδ and Nε atoms make van der Waals interactions with the AN6 and AC6 atoms of the NAD).

NAD	**C3stau2** (dist Å)		**C3bot1** (dist Å)	
nicotinamide	—		Phe-183-ring	
NO7	Leu-86-N Gln-140-NE2	3.0 3.1	Gly-129-N	3.0
NN7	Leu-86-O	3.2	Gly-129-O	2.9
NO2′	Glu-180-OE1	2.5	Glu-214-OE1 Gln-212-NE2	2.9 3.3
NO3′	Thr-37-O	2.9	Thr-80-O	2.7
NO1	—	—	Arg-128-NH1 Arg-128-NH2	3.3 3.3
NO2	Arg-150-NH2 Arg-150-NH1	2.8 3.1	Arg-186-NH1	3.0
AO1	Arg-85-NH2	3.2	Arg-128-NH1 Arg-128-NH2	3.3 2.8
AO2	Arg-48-NH1 Arg-48-NH2	2.9 2.8	Asn-87-ND2	2.8
AO3′	Asn-88-ND2	3.0	Asp-131-OD2	2.8
AO2′	Asn-88-N Asp-90-OD2	2.7 2.4	Asp-131-N Asp-130-OD1	2.8 2.7
AN6	Glu-133-OE1	2.9	Glu-169-OE1	2.9
AN7	—	—	—	—
adenine	—		Arg-91-side chain	
NAD	**Iota toxin** (dist Å)		**VIP2** (dist Å)	
nicotinamide	—		—	
NO7	Arg-296-N	2.7	Trp-350-N	2.6
NN7	Arg-296-O	2.9	Trp-350-O	3.0
NO2′	Glu-380-OE1 Glu-301-OE2	3.1 2.9	Glu-428-OE1	2.8
NO3′	—	—	—	—
NO1	Arg-352-NH1	2.8	—	—
NO2	Arg-352-NH2	2.6	—	—
AO1	Arg-295-NH1	3.3	—	—
AO2	—	—	—	—
AO3′	—	—	—	—
AO2′	Gln-300-NE2	3.1	—	—
AN6	Asn-335-OD1	3.2	—	—
AN7	Asn-335-ND2	3.1	—	—
adenine	—		Arg-315-side chain	

Table 4.4: Summary of the interactions of C3stau2, C3bot1, Iota toxin and VIP2 with NAD/NADH. Potential hydrogen bonding interactions for C3stau2 and VIP2 were calculated using HBPLUS (McDonald and Thornton 1994) with a distance cut-off of 3.3Å and D-H-A angles >90°.

Interactions for C3bot1 and the Iota toxin were taken from the crystal structure references. (Menetrey *et al.* 2002; Tsuge *et al.* 2003). The nicotinamide and adenine interactions are hydrophobic in nature.

have been reported for C3stau2 (Wilde *et al.* 2002c), C3bot1 (Menetrey *et al.* 2002), as well as for the Iota toxin (Sakurai *et al.* 2003; Tsuge *et al.* 2003) and confirm the importance of these arginines in NAD binding or hydrolysis. Surprisingly, comparison of the structures of all four enzymes show that the interactions between these arginines and NAD are not always identical (Figure 4.12 and Table 4.4). The two arginines conserved in the C3-like and binary toxins interact with the central phosphates in all the structures except VIP2. VIP2 does undergo a similar conformational change to C3stau2 upon NAD binding, and although it retains these arginines in positions similar to those of C3stau2, they are not close enough to provide the same type of interactions. This difference seems to correlate with the difference in the orientation of the NAD in the binding pocket of VIP2 and does not necessarily reflect the nature of transition state binding, which could be similar for all four enzymes. Alternatively, the need of the first serine in the STX motif for Iota toxin NAD hydrolysis (Nagahama *et al.* 2000), which is not the case in C3stau2 and C3bot1 (Menetrey *et al.* 2002; Wilde *et al.* 2002c), may indicate a difference in the mode of transition state stabilisation between the C3-like and binary toxins.

The arginine at position 3 is only conserved in the C3-like toxins but also occurs in the binary toxins VIP2 and C2 from *C. botulinum* and may be presumed to have a conserved function. This arginine has also been shown to be essential for NAD binding to C3stau2 (Wilde *et al.* 2002c) and is seen binding to the AO2 atom of NAD. However, in C3bot1, despite having the potential to bind to the central phosphate, it adopts a different conformation and its Cδ and Nε atoms engage in van der Waals interactions with the AN6 and AC6 atoms of the NAD. In VIP2, the guanidinium group makes π–π interactions with the adenine ring suggesting that this residue has several possible means of positioning the NAD, or that these orientations are preliminary to further change on hydrolysis. The variation in the orientation of this residue also coincides with the larger divergence of this (adenine) end of the binding pocket.

4.3.6 Implications for RhoA and RalA binding

Although structural data and mutational analysis have some way towards identifying residues involved in NAD cleavage and ADP-ribosylation amongst the C3-like toxins, only Tyr 175 in C3stau2 (Wilde *et al.* 2002c) and its equivalent, Phe 209 in C3bot1 (Wilde *et al.* 2002a), have been identified as essential for binding RhoA by

mutational analysis. The variation of this residue between the two enzymes may have some part to play in protein target specificity, although C3cer, which does not ribosylate RhoE (Wilde *et al.* 2003) also has a tyrosine at this position in the sequence alignment (Figure 2.3, Chapter 2). However, more detailed mutational analysis of the *C. difficile* Iota toxin suggests that helix α8 (equivalent to helix α3 in C3stau2) may also be partly responsible for the binding of protein substrate, which for the Iota toxin is actin (Tsuge *et al.* 2003). Also, two residues from this helix, Tyr 246 and Asn 255, in structurally equivalent positions to Ser 40 and Asn 44 in C3stau2, have been identified as residues important for protein substrate binding (Sakurai *et al.* 2003). Asn 255 was additionally identified as being important for protein substrate specificity (Sakurai *et al.* 2003). Interestingly, VIP2 and Iota toxin, which share actin as a substrate, align very well along helix 3 with an r.m.s. deviation of 0.3 Å over 12 Cα atoms, whereas C3stau2 and C3bot1, which have slightly

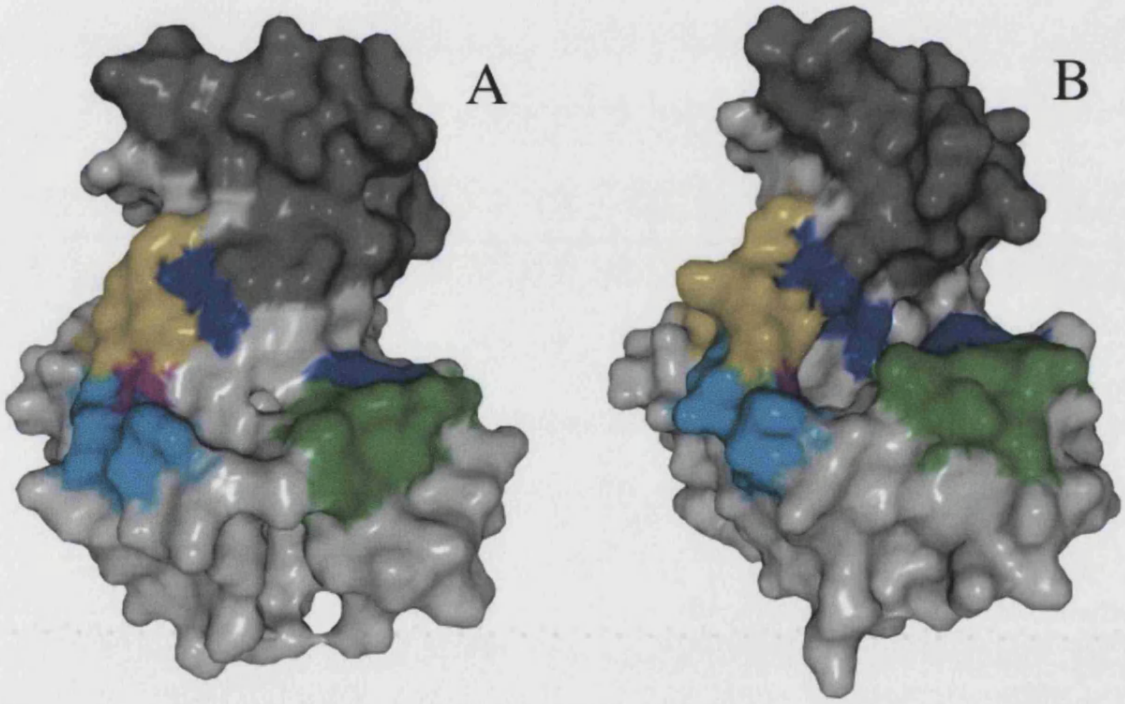


Figure 4.13: Surface diagrams of the NAD bound conformation of A, C3stau2 (PDB=10JZ) (Evans *et al.* 2003), and B, C3bot1 (PDB=1GZF) (Menetrey *et al.* 2002). The ARTT loop is shown in turquoise, the PN binding loop in yellow, the variable domain in grey and helix α3 in green. The NAD binding cleft is partially defined by N178/N212 (pink) and the three arginines which bind NAD (blue).

different substrate specificities, do not align so well and have an r.m.s. deviation of 1.2 Å. The N-terminal end of helix α3 in C3stau2 is shifted up to 4 Å away from the NAD binding site compared to the position of the corresponding helix in C3bot1. This is partly attributable to the inclusion of two prolines, Pro 42 and Pro 46, in this helix in C3stau2. By analogy with the Iota toxin, it is possible that specific residues in this helix are involved in Rho binding and that the change in conformation of this helix (Figure 4.13) contributes to the subtle difference in substrate specificity observed between C3bot1 and C3stau2. Further mutational analysis and/or the structure of a ternary complex will be required to substantiate this hypothesis.

Recently it has been reported that the ras-related GTPase RalA acts as an inhibitor of C3bot1, C3lim and C3cer, but not of C3stau2 (Wilde *et al.* 2002a). RalA is not ribosylated by these C3s but appears to bind them in a manner that blocks RalA functions, prevents C3 ribosylation of RhoA and enhances C3 NAD glycohydrolase activity. Mutation of Phe 209 in the C3bot1 ARTT loop did not decrease RalA binding, suggesting that RalA binds in a different way to RhoA. The level of inhibition of RhoA ribosylation varied from 90% inhibition of C3bot1, to none at all for C3stau2 (Wilde *et al.* 2002a). It is likely that this variation in inhibition, and therefore binding, is attributable to differences in structure. The structures of C3stau2 and C3bot1 reveal that the segment immediately following β1 not only varies in length but also adopts a different secondary structure. The variability of this region as well as its role in positioning residues for NAD hydrolysis make it a candidate to contribute to the RalA binding site.

4.3.7 Further Thoughts

The recent discovery that EDIN A also ribosylates RhoE/Rnd3 as well as C3stau2 (Wilde *et al.* 2003) suggests that this may be a common feature of the C3 exoenzymes from *S. aureus*. This reaction was not quantified and so how the differences in the sequence of these enzymes may effect the activity is not known. Comparison of the sequence of helix α3 between the *S. aureus* exoenzymes and the Clostridial exoenzymes does not reveal any residues that are substantially different between the two groups but conserved within them (Figure 2.3, Chapter 2). However, as discussed above, if this region is involved in substrate specificity it is probably due to the orientation of the helix. This

orientation may be in part due to the number and positions of prolines within it, but may also be positioned by other unidentified features in the surrounding sequence.

It is also interesting to note, that of the residues identified in RhoA as required for recognition of C3bot1, only three out of six, the three glutamates, are conserved in RhoE/Rnd3 (Figure 2.6, Chapter 2). Of the remaining three residues, Val 43 is not conserved in RhoC either and is replaced by an isoleucine. RhoE/Rnd3 has a serine at this position suggesting that a hydrophobic residue rather than a polar one is required for C3bot1 recognition. More significantly, Arg 5 and Lys 6, conserved in RhoA-C are replaced by a lysine and a cysteine in RhoE/Rnd3. It is possible that the loss of a positive charge at this position prevents binding to C3bot1. Figure 4.14 compares the surfaces and the residue positions of the Rho GTPases. Further mutational analysis would help clarify these points but it is likely that the precise interactions involved in the specificity of these reactions will be hard to unpick without a structure of a C3-Rho complex or NMR binding studies.

Publication

The work in this chapter is being published in the Journal of Biological Chemistry and is currently available as an e-publication ahead of print. Both structures have been submitted to the pdb and their codes have been referenced with this chapter in the figure legends.

Evans, H.R., Sutton, J.M., Holloway, D.E., Ayriss, J.A., Shone, C. C., Acharya, K. R. (2003)
The Structure of C3stau2 and its Complex with NAD.
J. Biol. Chem. 278: 45924 - 45930.

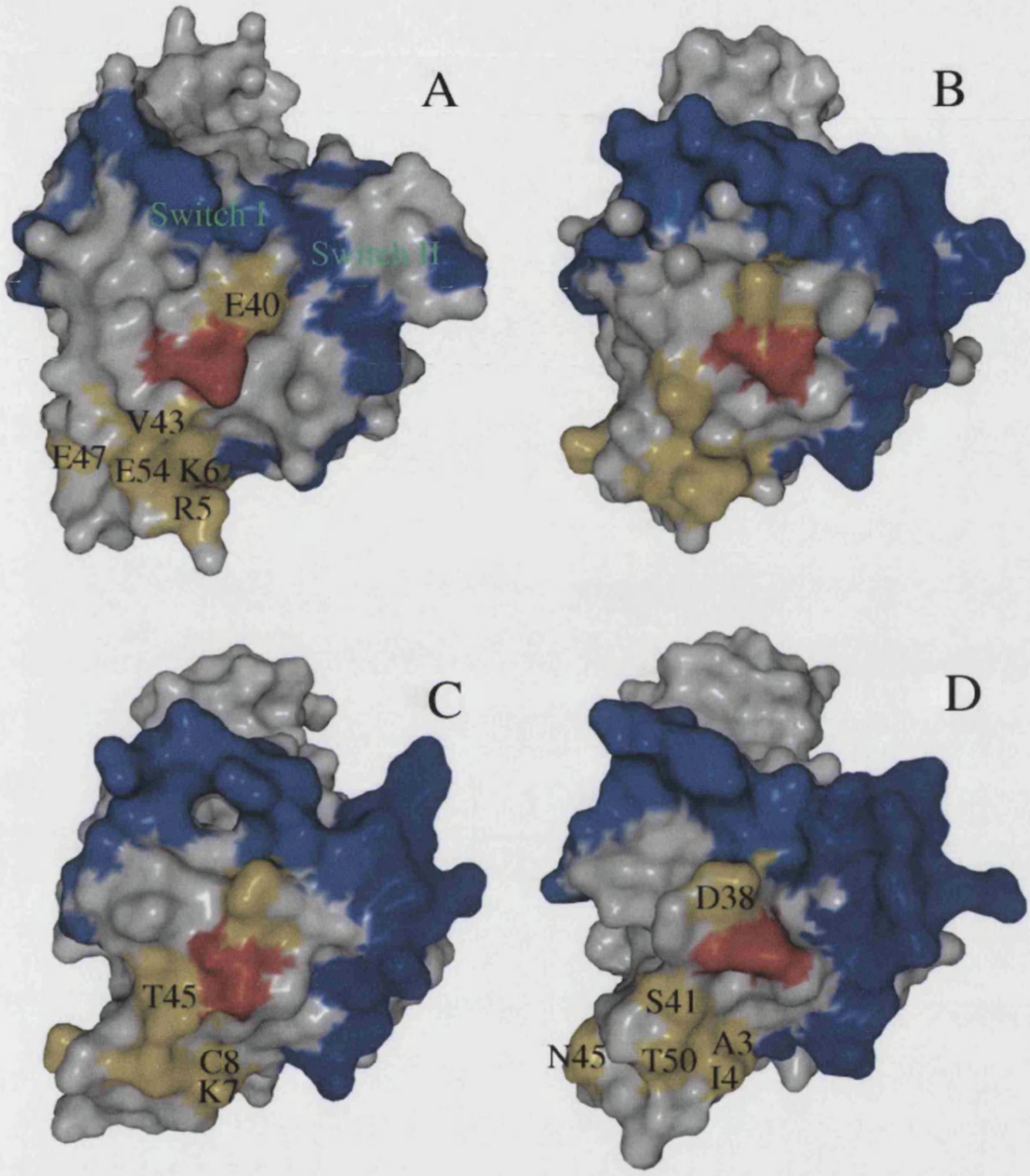


Figure 4.14: Surface comparison of the C3 binding residues in A, RhoA-GDP (PDB=1FTN) (Wei *et al.* 1997), B, RhoA-GTP (PDB=1A2B) (Ihara *et al.* 1998), C, RhoE (PDB=1M7B) (Fiegen *et al.* 2002) and Rac1 (PDB=1DS6) (Scheffzek *et al.* 2000; Evans *et al.* 2003). Switch I and II are in blue, N41 from RhoA and its equivalents in red and the residues shown to be essential for C3bot1 recognition of RhoA (and their equivalents) in yellow. Residues are labelled for RhoA and where they differ from RhoA in the other structures.

Chapter 5

Characterisation of C3bot1 and C3stau2

5.1 Introduction

Rho GTPases interact with many proteins including GAPs, GEFs and GDIs as well as their effector proteins. The specificity of these proteins for the various Rho GTPases varies from the highly specific effector proteins to the common rhoGDIs which bind all Rho GTPases. All of these proteins bind to Rho via its switch regions, and regulation of these interactions is mediated by the conformation of the switches depending on the guanine nucleotide bound. The specificity of these interactions, however, varies. For example, rhoGDI which binds all Rho GTPases but not Ras, relies on only two conserved residues, Arg66 in switch II and His103 (Grizot *et al.* 2001). Whereas, the GAP and GEF interactions, involving only a handful of residues, are more specific for particular Rho GTPases.

Although much is known about the specificity of Rho GTPases and their regulators and effectors, less is known about the specificity of C3bot1-RhoA interaction. It has always been presumed that the ribosylation of RhoA was a compulsory order mechanism with NAD binding first. Recent papers, however, have reported that C3bot1 binds RhoA in the absence of NAD. An assay where Rho-GST is used to 'pull down' C3bot1 in the absence of NAD has been used to characterise the specificity of the ribosylation (Wilde *et al.* 2000). This assay has subsequently been used to characterise the catalytic residues of C3stau2 (Wilde *et al.* 2002b; Wilde *et al.* 2002c) and C3cer (Wilde *et al.* 2003). The limited characterisation of the C3bot1-RhoA interaction with this assay identified the Rho residues Arg 4, Lys 6, Glu 40, Val 43, Glu 47, and Glu 54 as important (Wilde *et al.* 2000). By contrast, the only C3bot1 residue proposed to be involved in the interaction is Phe 209 (Han *et al.* 2001), for which the equivalent residues in C3stau2 (Y175) and C3cer (Y180) have been shown to be important for Rho binding by mutagenesis (Wilde *et al.* 2002c; Wilde *et al.* 2003).

Although the specificity of C3bot1 for RhoA-C and not other Rho GTPases is well established, it was recently shown that C3bot1 binds, but does not ADP-ribosylate, the ras-related GTPase RalA. This interaction blocks RalA functions, prevents C3 ribosylation of RhoA and enhances C3 NAD glycohydrolase activity (Wilde *et al.* 2002a). Glutathione pull down assays also demonstrated this interaction for C3lim and C3cer and the affinity of

RalA for C3lim was shown to be 12nM by fluorescence titration (Wilde *et al.* 2002a). Interestingly, C3stau2 did not appear to bind RalA, suggesting that RalA binds C3 exoenzymes in a manner different to RhoA, raising another interesting question of specificity.

This chapter describes the basic characterisation of our recombinantly produced C3bot1 and C3stau2. To investigate the specificity of C3bot1 for RhoA, binding experiments in the presence and absence of substrate analogues were performed to attempt to stabilise a complex between C3bot1 and RhoA for crystallisation. A limited characterisation of the C3bot1-RalA was also attempted.

5.2 Methods

Purification of RhoA and RalA- For assaying the activity of C3bot1 and C3stau2 a supply of RhoA and RalA was needed. These were made recombinantly as GST-fusion proteins. *E. coli* cell paste containing soluble, overexpressed RhoA-GST or RalA-GST was obtained from Dr. J.M. Sutton at the HPA. 100ml cultures in Terrific broth with 100µg/ml ampicillin and 0.5% glucose were grown overnight at 30°C. 20ml of these cultures were transferred to a litre culture of the same media and grown at 30°C. Protein production was induced at 25°C with 1mM (final concentration) IPTG when the cells reached a density of 0.8 absorbance units. Batches of 500mls cells were harvested, resuspended in lysis buffer (50mM Hepes pH7.3, 0.5M NaCl, 5mM $MgCl_2$, 1mM DTT and 1mM PMSF) and stored at -20°C. Thawed cells were lysed by sonication on ice (5x 40s bursts at 26microns – Soniprep 150) and cell debris were separated by centrifugation. The GST fusion protein bound a GST-trap column (Amersham) equilibrated with thrombin resuspension buffer (50mM Hepes pH7.3, 150mM NaCl, 5mM $MgCl_2$, 2.5mM $CaCl_2$, and 1mM DTT) and was eluted with the same buffer plus 20mM reduced glutathione. The GST was cleaved at 4°C for 4-6 days with 4U thrombin/mg fusion protein. The cleaved GST re-bound the GST column and the non-bound protein was concentrated in a centrifugal 15KDa cut off concentrator (Amicon). Finally, the protein was purified in batches of 300µl on a Superdex-200 HR 10/30 gel filtration column run at 0.75ml/min. RhoA was purified in 20mM Hepes pH7.3, 20mM NaCl, 5mM $MgCl_2$ and 1mM EDTA, concentrated to 10-12mg/ml, aliquoted and stored at -70°C. RalA was purified in 20mM Hepes pH7.3, 120mM NaCl, concentrated to 4mg/ml, aliquoted and stored at -70°C.

ADP-ribosylation assay for C3bot1- 50µl 0.01-2.5µg/ml C3bot1 was mixed with a reaction mix containing 10µl DTT assay buffer (50mM Hepes/NaOH pH 7.5, 5mM $MgCl_2$ and 5mM DTT), 20µl 25µg/ml Rho A in assay buffer and 20µl 4µM NAD with 50µCi/ml [*adenylate-* P^{32}] NAD. The reaction mix was incubated at 37°C for 2hrs and stopped with 100µl 2mg/ml BSA followed by 0.5ml 24% trichloroacetic acid (TCA). The precipitated protein was collected by high speed microfuge centrifugation and the pellets washed with 1ml 12% TCA. The pellets were resuspended in 1ml Hepes/NaOH pH 7.5. 5mls of scintillation fluid was added to each sample before counting.

ADP-ribosylation assay for C3stau2- 50µl 0.5-500µg/ml C3stau2 (in 20mM Hepes pH 7.4, 0.1mg/ml BSA) was mixed with a reaction mix containing 10µl assay buffer (50mM Hepes/NaOH pH 7.4, 5mM $MgCl_2$), 20µl 2mg/ml RhoA in assay buffer and 20µl 200µM NAD with 50µCi/ml [*adenylate-* P^{32}] NAD. The reaction mix was incubated at 37°C for 1hr and stopped with 100µl 2mg/ml BSA followed by 0.5ml 24% TCA. The precipitated protein was collected by high speed microfuge centrifugation and the pellets washed with 1ml 12% TCA. The pellets were resuspended in 1ml Hepes/NaOH pH 7.4. 5mls of scintillation fluid was added to each sample before counting.

Inhibition assay- 20µl 70ng/ml C3bot1 was added to a reaction mix containing 20µl DTT assay buffer (50mM Hepes/NaOH pH 7.5, 5mM $MgCl_2$ and 5mM DTT), 20µl 25µg/ml Rho A in assay buffer, 20µl 1mM NAD analogue and 20µl 50µCi/ml [*adenylate-* P^{32}] NAD. The reaction mix was incubated, stopped and counted as above. The substrate analogues used are shown in Table 5.1. The C3bot1 used for these experiments was bought from Sigma and the RhoA was provided by Dr. C.C. Shone at the HPA.

Assessment of analogues as inhibitors of substrates- As inhibition assay, but samples were pre-incubated for 1hr at 37°C before addition of the [*adenylate-* P^{32}] NAD either with, or without, substrate analogues 1, 5, 6 and 9 (Table 5.1). The stock concentrations of the substrate analogues 1,5, 6 and 9 used were 0.5mM, 0.5mM, 0.2mM and 1mM respectively. The C3bot1 used for these experiments was bought from Sigma and the RhoA was provided by Dr. C.C. Shone at the HPA.

NAD glycohydrolase assay for C3stau2 - 50µl 500µg/ml C3stau2 in 20mM Hepes/NaOH pH7.4, 0.1mg/ml BSA, was mixed with 50µl 200µM NAD including 20µM [nicotinamide- ^{14}C] NAD in Hepes/NaOH pH 7.4. Samples were incubated over a 5hr time course and stopped by vortexing with 1ml water-saturated ethyl acetate for 3x 10 seconds. 0.75mls of the ethyl acetate phase was added to 5mls scintillation fluid and counted.

NAD glycohydrolase assay for C3bot1 - 50µl 250µg/ml C3bot1 in 20mM Na.Hepes pH7.4, 0.1M NaCl, 5mM $MgCl_2$, 0.1mg/ml BSA, was mixed with 50µl 200µM NAD including 20µM [nicotinamide- ^{14}C] NAD in Na.Hepes pH 7.4, 0.1M NaCl, 5mM $MgCl_2$. Samples were incubated over a 5hr time course and stopped by vortexing with 1ml water-

saturated ethyl acetate for 3x 10 seconds. 0.75mls of the ethyl acetate phase was added to 5mls scintillation fluid and counted.

Glutathione-captured RhoA-GST binding assays- 1mg C3bot1 and 0.5mg RhoA-GST in 5mls binding buffer (50mM Na.Hepes pH7.3, 2mM $MgCl_2$, 1mM EDTA, 1mM DTT), were incubated overnight at 4°C. The mixture was loaded onto a GST-Trap column (Amersham) in chilled binding buffer and bound material eluted with binding buffer plus 20mM glutathione.

Gel filtration binding assays- Either 60µg C3 and 30µg RhoA, or 60µg RhoA and 30µg C3 were incubated overnight at 4°C in 50-75µl 50mM Na.Hepes pH7.3, 0.1M NaCl, 5mM $MgCl_2$ and 1mM DTT. The proteins were separated by gel filtration with a Superdex-200 HR 10/30 or Zorbax Bio Series GF-250 gel filtration column using the same buffer. Fractions were precipitated by saturated TCA and viewed by SDS-PAGE. Assays were also carried out in the presence of 150µM 3-acetylpyridine adenine dinucleotide. A similar assay was used for RalA.

Growth of NIH 3T3 cells- The cells were cultured with the help of Dr J. Ayriss at the HPA. NIH 3T3 cells were grown at 37°C with 5% CO_2 in Dulbecco Modified Eagle Medium (DMEM) with 10% Calf serum, 2mM L-Glutamine and 0.1mg/ml Penicillin/Streptomycin. Cells were sub-cultured every 3-4 days.

Visualising changes in cell morphology- Cells were seeded on coverslips or chamber 8 well slides at a density of $2x10^4$ or $2x10^5$ and grown overnight at 37°C at 5% CO_2. For the assay, the growth media was replaced with uptake media containing DMEM, 0.5M sucrose, 10% PEG1000, 0.1mg/ml ovalbumin and 15-150µg/ml exoenzyme, or uptake media for the control. The cells were incubated for 8mins at 37°C before washing twice with growth media. The cells were then incubated for 1-3hrs in fresh growth media and visualised with an inverted LEICA DMIRB microscope at 5x magnification.

Visualising actin changes with Rhodamine-phalloidin- Cells were seeded on coverslips at a density of $2x10^4$ or $2x10^5$ and grown overnight at 37°C at 5% CO_2. The assay was performed as above. To fix the cells, the coverslips were washed twice in 1x

PBS, pH 7.4 and incubated with 4% paraformaldehyde in PBS for 10mins at room temperature. The cells were permeabilised by incubation with 0.1% Triton X-100 in PBS for 4mins and blocked by incubation with 1% bovine serum albumin (BSA) for 20-30mins. The cells were stained for 20mins with 0.4μM Rhodamine-phalloidin in PBS containing 1% BSA and mounted on microscope slides. The cells were visualised with an inverted LEICA DMIRB confocal microscope with water emersion at 63x magnification.

Amine coupling of chips for Biacore- Amine coupling was performed following the manufacturers instructions. The chip surface was activated by injection of 35μl 1:1 mixture of NHS (0.05M N-hydroxysuccinimide) and EDC (0.2M N-ethyl-N'-(dimethylaminopropyl)-carbodiimide). 70μl 200μg/ml C3bot1 or 35μl 100μg/ml RhoA-GST in 10mM Acetate pH 4 were injected over the activated surface of flow cell 1 with HBS-EP buffer (10mM Na.Hepes pH7.4, 150mM NaCl, 3mM EDTA, 0.005% Surfactant P20) at 10μl/min. The remaining amide groups were deactivated by injection of 35μl 1M ethanolamine hydrochloride pH8.5. The coupling was repeated for flow cell two using buffer instead of protein to create a control. NHS, EDC, HBS-EP and ethanolamine hydrochloride were all supplied by Biacore. The RhoA-GST for these experiments was supplied by Dr. J.M. Sutton at the HPA.

Injection of Biacore samples- All experiments were performed using the recommended HBS-EP running buffer (10mM Na.Hepes pH7.4, 150mM NaCl, 3mM EDTA, 0.005% Surfactant P20) at speeds of 20-80μl/min. Protein samples were exchanged into this buffer using buffer exchange columns (PD10, Amersham). 100μl samples were injected using the recommended technique. Washing of the injection loop was delayed 600s to prevent interference with the trace. Both chips were washed using 10μl 50-200mM sodium carbonate pH 9.6. The coupling was repeated for flow cell two using buffer instead of protein to create a control. HBS-EP was supplied by Biacore. The RhoA-GST and RalA-GST for these experiments was supplied by Dr. J.M. Sutton at the HPA.

5.3 Results and discussion

5.3.1 Purification of RhoA and RalA

The investigation of C3bot1 and C3stau2 binding to either RhoA or RalA, required expression and purification of both these proteins. Also, it was hoped that either a complex of C3bot1-RhoA-substrate analogue, or C3bot1-RalA could be produced in large amounts for crystallisation.

5.3.1a Results

RhoA is routinely expressed as a GST-fusion protein and *E. coli* cultures containing overexpressed RhoA-GST were provided by Dr. J. M. Sutton at the HPA. Purification of the fusion protein on a glutathione column produced a reasonably clean peak with a yield of 30mgs fusion protein per litre cells. Initial trials indicated that the cleavage reaction required 0.2U thrombin to cleave 20μg fusion protein at room temperature in 16hrs. However, protein purified in this way was shown to be inactive and was not ADP-ribosylated by C3bot1 in assays, whereas the un-cleaved protein could be ribosylated. The cleavage reaction was eventually refined to 4U thrombin/mg fusion protein at 4°C for 4-6 days to prevent denaturing. The RhoA was separated from the GST which re-bound the glutathione column (Figure 5.1). More latterly, RhoA was also purified

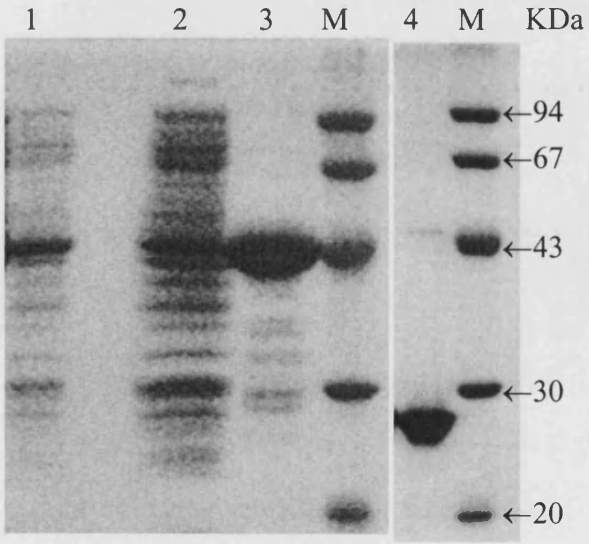


Figure 5.1: 12% SDS PAGE showing the purification of RhoA. Lane 1, the soluble cell extract, 2, the protein which did not bind the GST-trap column (Amersham), 3, the peak eluted from the GST-trap column, 4 the cleaved protein after the second GST-trap purification and M, the low molecular weight marker (Amersham).

by gel filtration to prevent any impurities confusing the gel filtration binding assay. After concentration, the RhoA was stored at -70°C as it was found to precipitate during prolonged storage at 4°C.

RalA-GST was also provided overexpressed in *E. coli* cultures by Dr. J. M. Sutton at the HPA. The expression and purification of this protein were carried out as for RhoA and not optimised. A significant proportion of the fusion protein was insoluble after cell lysis by sonication and a yield of only 15mgs fusion protein resulted from 1litre of cells. RalA-GST also proved to be quite resistant to cleavage and incubation with 7U thrombin/mg fusion protein for four nights at 4°C did not result in total cleavage (Figure 5.2). Further purification to remove the GST by glutathione column did not result in pure protein as judged by SDS-PAGE so the RalA was further purified by gel filtration. Unfortunately this resulted in a yield of only 0.39mg per litre cells. The pure protein was concentrated to 4mg/ml and stored at -70°C.

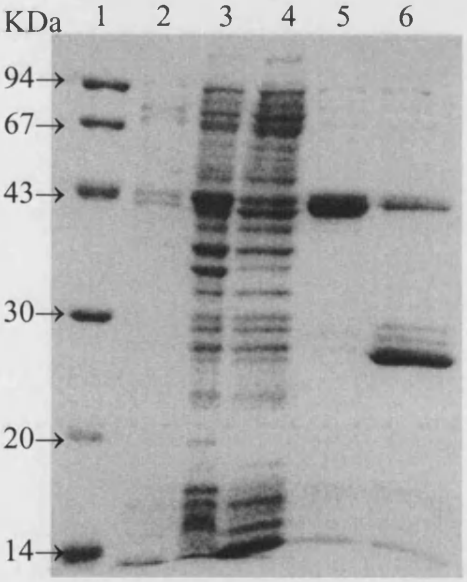


Figure 5.2: 12% SDS PAGE showing the purification of RalA. Lane 1, the low molecular weight marker (Amersham), 2, the soluble cell extract, 3, the insoluble extract, 4, the protein that did not bind the GST-trap column (Amersham), 5, the peak eluted from the GST-trap column and 6, the cleaved protein after incubation with 7U thrombin/mg fusion protein for four nights at 4°C.

5.3.1b Discussion

RhoA has previously been expressed with a GST-tag (Wei *et al.* 1997), a His tag (Shimizu *et al.* 2000) and a FLAG affinity tag (Longenecker *et al.* 1999) for crystallisation. However, the yield of these purifications was not commented on except for the FLAG tag. RhoA-FLAG, was expressed in yeast and purified as a complex with rhoGDI. Therefore, the resulting yield of 4-6mg from 18 litres of cells (Longenecker *et al.* 1999) is not comparable. Although expression as a GST-fusion protein gives an initial yield equivalent to 15mgs of RhoA, the inefficiency of the cleavage reaction make this method less convenient for very large scale purification. A trial to assess whether the cleavage reaction is more efficient when RhoA is cleaved straight from the GST-beads in a batch method, showed that the cleavage was still very inefficient. This method was also hindered by the inefficient binding of RhoA-GST to the beads. After cleavage, purification by the GST-trap and gel filtration columns did yield very pure protein, although care was needed to prevent it precipitating upon concentration. Although no accurate yield measurements were made, I estimate that up to 4-6mgs highly pure protein could be purified from the initial 15mg with careful concentration, although this purification would take two weeks. The published method of purification for crystallisation using the GST-tag (Wei *et al.* 1997) cleaved the protein on the column which may have been more effective. Also, 10% PEG 400 was included in the buffer which may have helped the stability of the protein, although this was not commented on.

The poor cleavage of the GST from the RalA suggest that another construct may be more appropriate for purification. RalA has recently been purified using a His tag without cleavage for NMR titration experiments (Mott *et al.* 2003) but the purification details and yield were not reported. Although expression of RalA as a GST fusion protein results in a large proportion of the protein being insoluble, the yield from the purification could probably be improved by using more cleavage enzyme and maybe an anion exchange step rather than the gel filtration. Alternatively, the expression and lysis protocols could be optimised to produce more soluble protein. Also, more work would need to be done to characterise the activity of the RalA, probably using a radioactive GTPase assay. Unfortunately, the lack of success with the initial RalA experiments prevented continuation within the time frame of this Ph. D. project.

5.3.2 Characterisation of C3bot1

The following assays were carried out in order to optimise them for the inhibitor studies and to assess the activity of C3bot1 in comparison with C3stau2.

5.4.2a ADP-ribosyltransferase activity

The ADP-ribosyltransferase activity was measured by the incorporation of labelled ADP-ribose into RhoA. C3bot1 was incubated with RhoA and [*adenylate-* ^{32}P-adenylate] NAD for 1hr at 37°C and the labelled protein precipitated with TCA. The precipitated protein was resuspended with buffer and the radioactivity counted. Most of these assays were only done in duplicate as they were just intended to be comparative. These comparative assays showed the activity of our recombinant C3bot1 to be greater than commercially bought C3bot1 (Sigma). The method described in section 5.2 details the experiment performed to estimate the specific activity of C3bot1. Excess cold NAD was added to ensure the both substrates were in excess. However, plotting the number of moles ADP-ribosylated RhoA against C3bot1 concentration resulted in a curve rather than a straight line so the specific activity of 0.38mol ADP-ribose/mol enzyme/min could only be estimated from the beginning of this curve (Figure 5.3).

5.3.2b NAD glycohydrolase activity

In the absence of a protein substrate, C3 exoenzymes can catalyse the hydrolysis of NAD at a slower rate. This assay, based on an assay designed for diphtheria toxin (Kandel *et al.* 1974), was also adapted to include excess cold NAD. 100µl samples containing 5.3µM C3bot1 and 100µM NAD (including 10µM [*nicotinamide-* ^{14}C] NAD) were incubated at 37°C over a time course of 5hrs. Reactions were stopped by mixing with 1ml water-saturated ethyl acetate. 750µl of the ethyl acetate phase, which captures >95% of the nicotinamide (Kandel *et al.* 1974), was mixed with 5mls scintillation fluid for counting. Plotting of the number of moles of nicotinamide released against time shows a linear relationship for the first 2hrs (Figure 5.4) from which a specific activity of 0.12 mol/mol enzyme/min can be calculated.

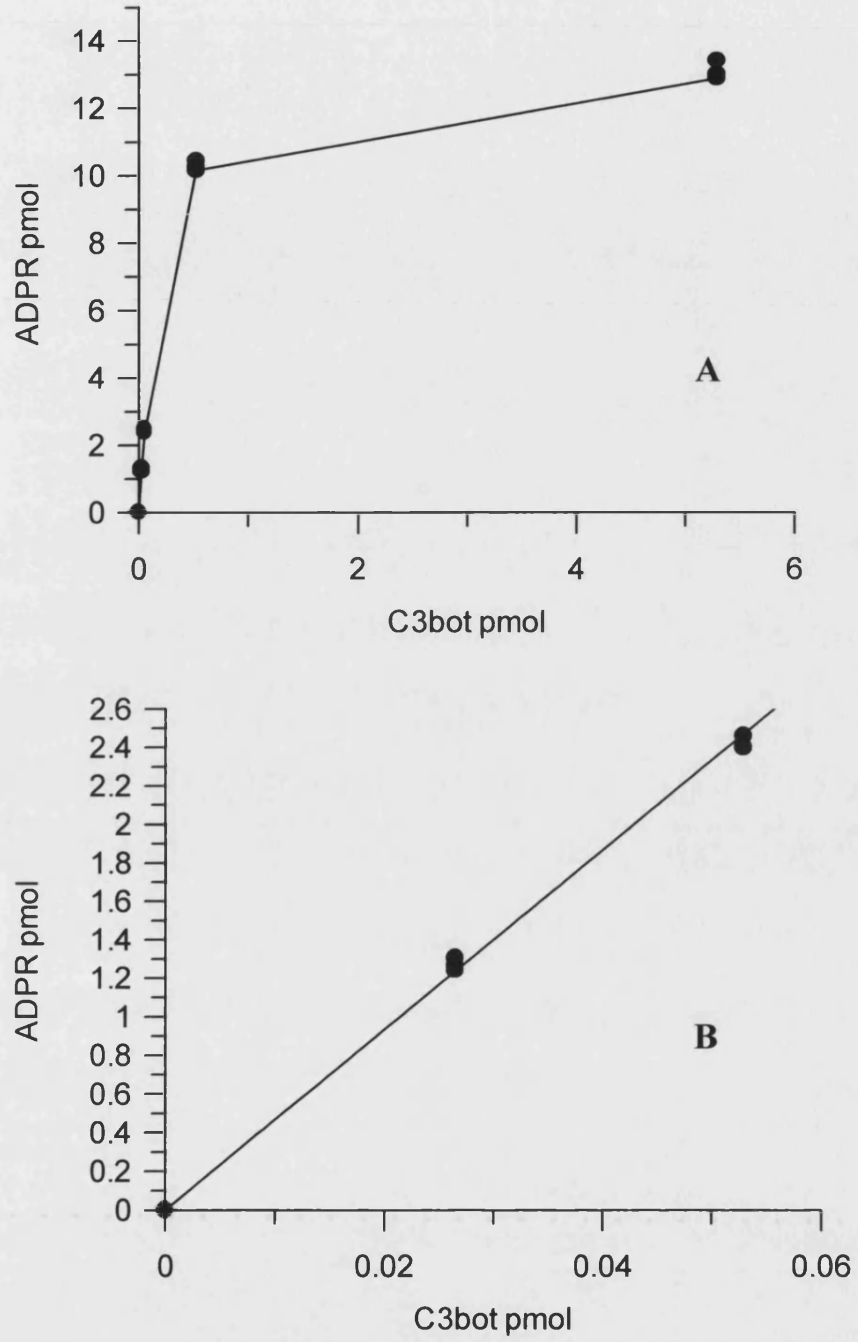


Figure 5.3: The transfer of [*adenylate-* ^{32}P-adenylate] ADP-ribose to RhoA over 2hrs at different C3bot1 concentrations. A, the relationship is not linear showing that the substrates may be not in excess for the higher concentrations of C3bot1. **B,** the first part of the curve in A is linear and can be used to calculate specific activity. The gradient is 46.7pmol ADP-ribose/pmol C3bot1 from which a specific activity of 0.38mol ADP-ribose/mol enzyme/min can be calculated. Assays were performed in triplicate and analysed using GraFit (Leatherbarrow 2001).

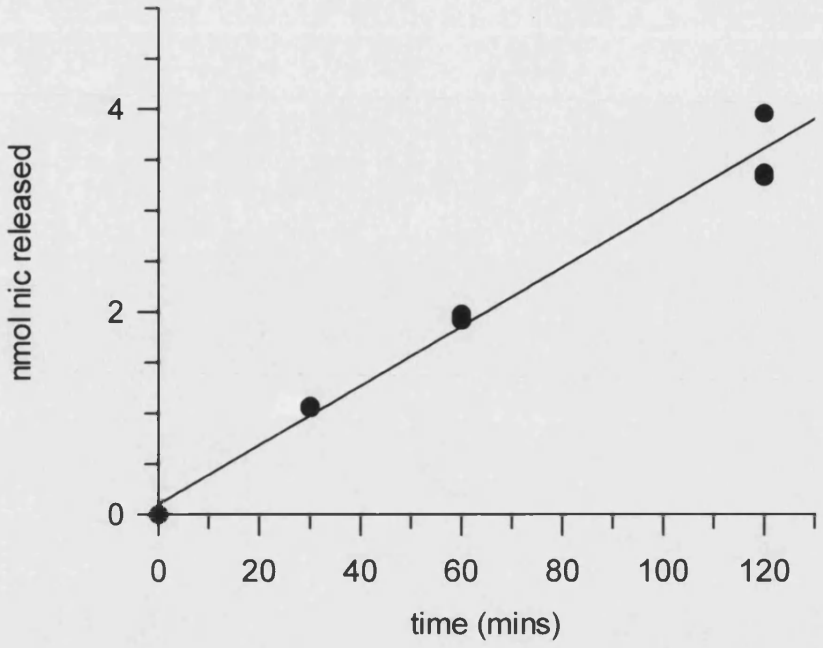


Figure 5.4: C3bot1 dependant release of nicotinamide from NAD over time. Triplicate 5.3μM C3bot1 samples were incubated with 200μM NAD including 20μM [*nicotinamide-* ^{14}C] NAD. The relationship is linear over 2hrs with a gradient of 0.03nmol/min, from which a specific activity of 0.12mol/mol enzyme/min can be calculated. Fitting performed with GraFit (Leatherbarrow 2001).

5.3.2c Cell assays

C3bot1 was initially characterised by its rounding effect on NIH 3T3 cells (Rubin *et al.* 1988) which was later attributed to the dissolution of actin filaments by inactivation of Rho proteins (Paterson *et al.* 1990). Therefore, the effect of our C3bot1 on cultures NIH 3T3 cells was assessed by visualisation of morphological changes and rhodamine-phalloidin staining of the F-actin filaments. As C3bot1 is not easily taken up into cells, an osmotic shock method (Rubin *et al.* 1988) was used to aid uptake. 1hr after the addition of the lowest concentration (15μg/ml) of C3bot1, reversible cell rounding (Figure 5.5) and actin dissolution were observed. Assays were performed in triplicate.

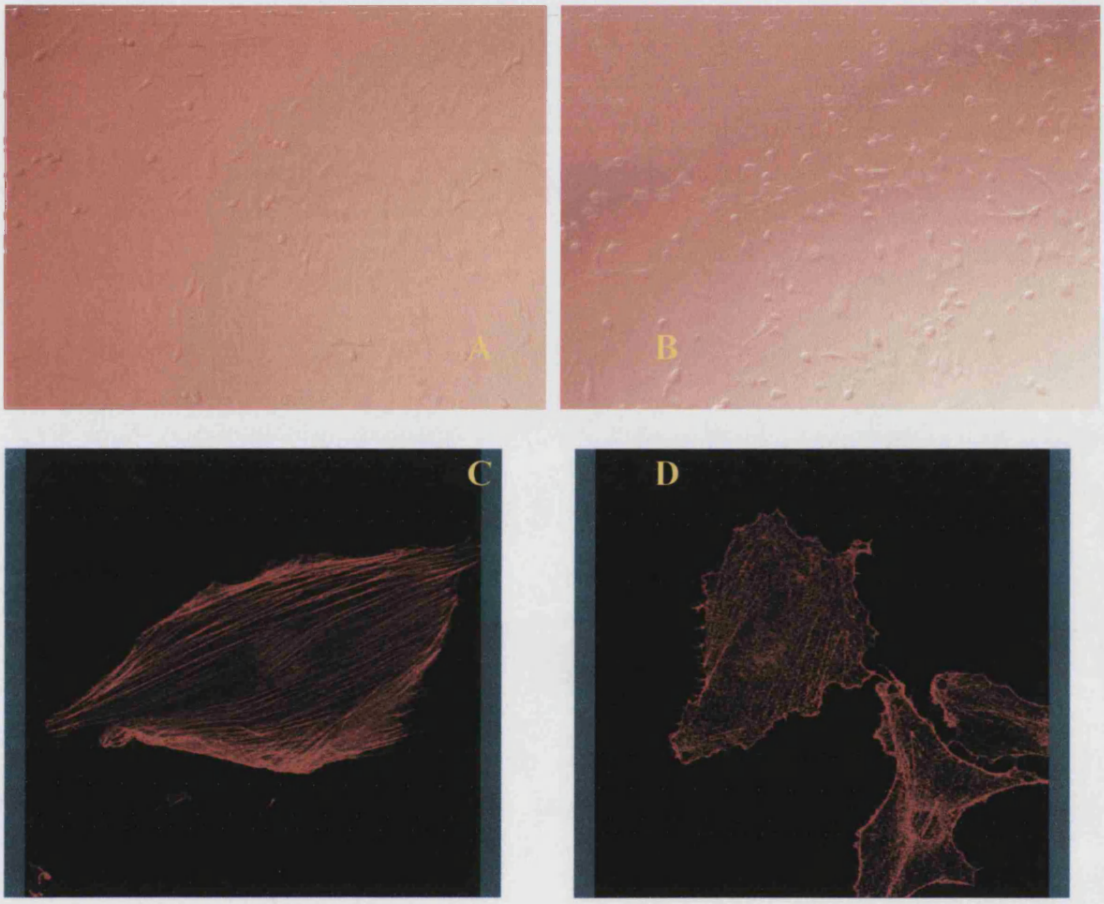


Figure 5.5: NIH 3T3 cell rounding upon addition of C3bot1. A and **C**: Controls without C3bot1, **B** and **D**: after incubation for 1-3hrs with 15µg/ml C3bot1. A and B were visualised on a LEICA DMIRB microscope at 5x magnification. C and D were visualised with an inverted LEICA DMIRB confocal microscope with water emersion at 63x magnification after staining with Rhodamine-phalloidin.

5.3.2d Discussion

The original use of the ADP-ribosylation assays was purely comparative for evaluation of NAD analogues. Early observations showed that our C3bot1 was at least as active, and often more so, that commercially bought enzyme. Later, we wanted to calculate the specific activity of C3bot1 for comparison with C3stau2 and other C3 exoenzymes. However, the level of activity observed, as judged by the counts of the scintillation counter, appeared to have decreased. This may have been due to the change in scintillation counter or a decrease in the activity of the C3bot1. However, it proved difficult to reproduce the previous results for more careful analysis or to re-optimise the assay to get a suitable figures for analysis. The current calculation of the specific activity, from the linear part of the graph, and taking into account the efficiency of the scintillation counter, does not seem to be as high as I would expect. Confusingly, comparison of the number of moles of RhoA ribosylated, and the original concentration of RhoA and NAD, shows that both substrates should still be in excess for all concentrations of C3bot1 by the end of the experiment. A non-linear relationship between enzyme concentration and product release under these conditions would suggest product inhibition of the reaction. The other alternative is that only a small percentage of the RhoA is active. However, further experiments would be required to characterise this reaction fully.

The consolation, on analysis of previously published activities for C3 exoenzymes (Table 5.1), is that there seems to be little agreement on its activity suggesting that either C3 exoenzymes vary considerably in activity between batches, or that the methods of calculation are not very accurate. Interestingly, all these activities are published from the same group (Aktories *et al.*) and references 3 and 4 (Table 5.1) are from the same author and still disagree. The published assays were all monitored by phosphorimaging of SDS-PAGE whereas ours were monitored by scintillation counting of resuspended precipitated protein. It is possible that the sensitivity of the assays varies. The lack of consistency between the published results and the fact that they all came from the same group, makes it hard to draw any concrete conclusions about the activity of this enzyme *in vitro*.

The NAD glycohydrolase activity for our C3bot1, 0.12mol/mol enzyme/min, is more comparable with published results for C3stau2 (0.61mol/mol/min) (Wilde *et al.* 2002c) and C3cer (0.29mol/mol/min) (Wilde *et al.* 2003) suggesting that these enzymes

have similar levels of NAD glycohydrolase activity, whatever the differences reported in transferase activity (Wilde *et al.* 2003). This would be expected from their common active site architecture. The effect of C3bot1 on NIH 3T3 cells also mirrored the published results (Rubin *et al.* 1988) with similar concentrations of the enzyme. In the light of these results, it seems even more confusing that there should be such discrepancy in the ADP-ribosyltransferase assays. Considering the stability of C3bot1 and the consistency of the other results it is likely that variations in the RhoA used for these assays maybe affecting the results. RhoA-GDP is a better substrate than RhoA-GTP for C3bot1 due to the conformation of the switches. Therefore it is possible that heterogeneity of the RhoA conformation either due to nucleotide binding or slight denaturation may affect the outcome of these assays. This may account for the inconsistencies in our results as the RhoA samples used varied in age (older for the later experiments) and some of the RhoA samples were provided by Dr. C.C. Shone at the HPA. An independent assay for RhoA activity may help validate future experiments in this area.

C3 type	Reference	Published rate	Specific activity (mol/mol/min)
C3bot1	1	6.4 nmol/mg/min	0.151
C3bot1	4	n/a	60
C3lim	1	3.1 nmol/mg/min	0.073
C3lim	2	12nmol/mg/min	0.28
C3lim	4	n/a	90
C3stau2	3	n/a	159
C3stau2	4	n/a	100
C3cer	4	n/a	22.25

Table 5.1: Reported specific activities for RhoA ADP-ribosylation by C3 exoenzymes. References are 1, (Just *et al.* 1992), 2, (Bohmer *et al.* 1996), 3, (Wilde *et al.* 2002c) and 4, (Wilde *et al.* 2003). All reported activities are calculated from phosphorimaging of SDS-PAGE. For conversion to mol/mol/min, 1mg C3 exoenzyme (~23.6 kDa) is taken to be approximately 42.4nmols.

5.3.3 Characterisation of C3stau2

These experiments were carried out to characterise C3stau2.

5.3.3a ADP-ribosyltransferase activity

C3stau2 was also assayed for its ADP-ribosyltransferase activity. Initial experiments using the same method as C3bot1 showed its activity to be much reduced to levels not detectable by the assay. It was also noted that whilst for C3bot1 the K_d for NAD is reported as 6μM (Chavan *et al.* 1992) and the K_m as 0.4μM (Just *et al.* 1992), for C3stau2 the K_d is reported as 20μM by fluorescence quenching assays (Wilde *et al.* 2002c). Therefore, the ribosylation assay was adapted with increased enzyme concentrations to raise the production of labelled ADP-ribose above the radioactive background and with increased substrate concentrations, including cold NAD, to keep the substrate in excess. The ten-fold dilutions of C3stau2 from the highest concentration of

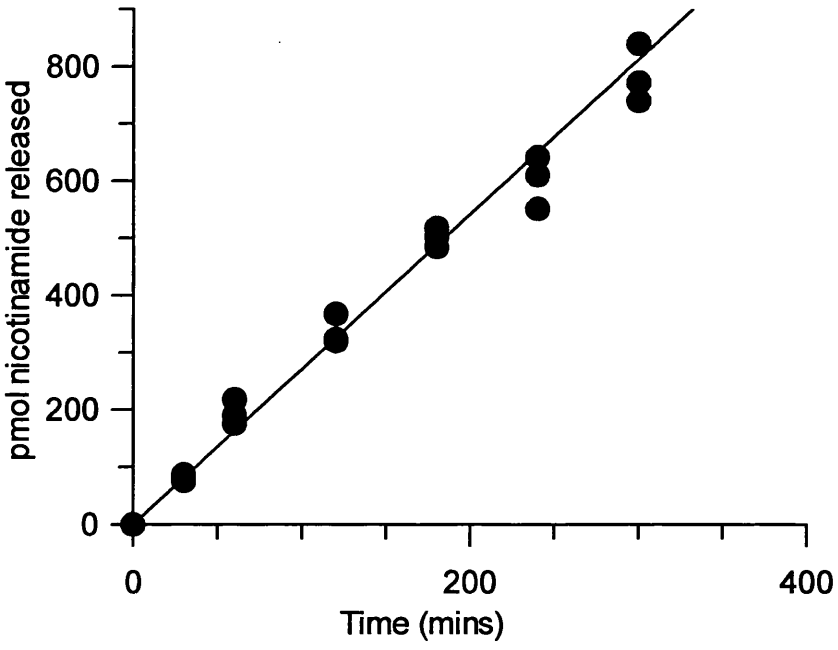


Figure 5.7: Linear increase of nicotinamide release with time for the NAD glycohydrolase activity of C3stau2. 10.5μM C3stau2 was incubated with 200μM NAD for up to 300mins. Triplicate samples were stopped by mixing with water-saturated ethyl acetate and counted. Results displayed are calculated taking into account the radioactivity of the NAD. Linear fitting with GrafFit (Leatherbarrow 2001) gives a gradient equal to 2.7 pmol/min giving an approximate specific activity of 0.005mol/mol enzyme/min.

10.6μM, incubated at 37°C for 1hr with 20μM NAD, including 10nM [*adenylate-* ^{32}P] NAD, and 19.4μM RhoA, did not show a directly proportional increase in production of labelled substrate with increase in enzyme. This suggested that the substrate may be limiting. However, an approximate specific activity of 0.19 mol ADPR/mol enzyme/min was calculated from the highest concentration of C3stau2 used. This is probably underestimated however as the NAD concentration for the assay should ideally be 10x the K_d and the RhoA was only 2-fold in excess.

5.3.3b NAD glycohydrolase activity

The NAD glycohydrolase activity of C3stau2 was also measured to further characterise the enzyme. 100μl samples containing 10.6μM C3stau2 and 100μM NAD (including 10μM [*nicotinamide-* ^{14}C] NAD) were incubated at 37°C over a time course of 5hrs. Reactions were stopped by mixing with 1ml water-saturated ethyl acetate. 750μl of the ethyl acetate phase was mixed with 5mls scintillation fluid for counting. Plotting of the number of moles of nicotinamide released against time shows a linear relationship (Figure 5.7) from which a specific activity of 0.005 mol/mol enzyme/min can be calculated.

5.3.3c Cell assays

Published characterisation of C3stau2 also reports similar cell rounding effects of NIH 3T3 cells upon addition of the enzyme, engineered to be taken up as a binary toxin (Wilde *et al.* 2001). It was hoped that cell rounding would therefore be observed by the addition of C3stau2 via the osmotic shock method used for C3bot1. NIH 3T3 cells were incubated with concentrations of up to 400μg/ml without visible changes in cell morphology. Staining with Rhodamine-phalloidin also produced no quantifiable difference in F-actin staining between C3stau2 incubated cells and controls, although the staining did appear lighter in some of the C3stau2 samples (Figure 5.8) All assays were done in triplicate.

5.3.3d Discussion

The differences seen between the published ADP-ribosyltransferase activity of C3stau2 (Wilde *et al.* 2001; Wilde *et al.* 2002c) and our activity may be in part due to differences in the methods used as discussed previously. The inadequate measurement of

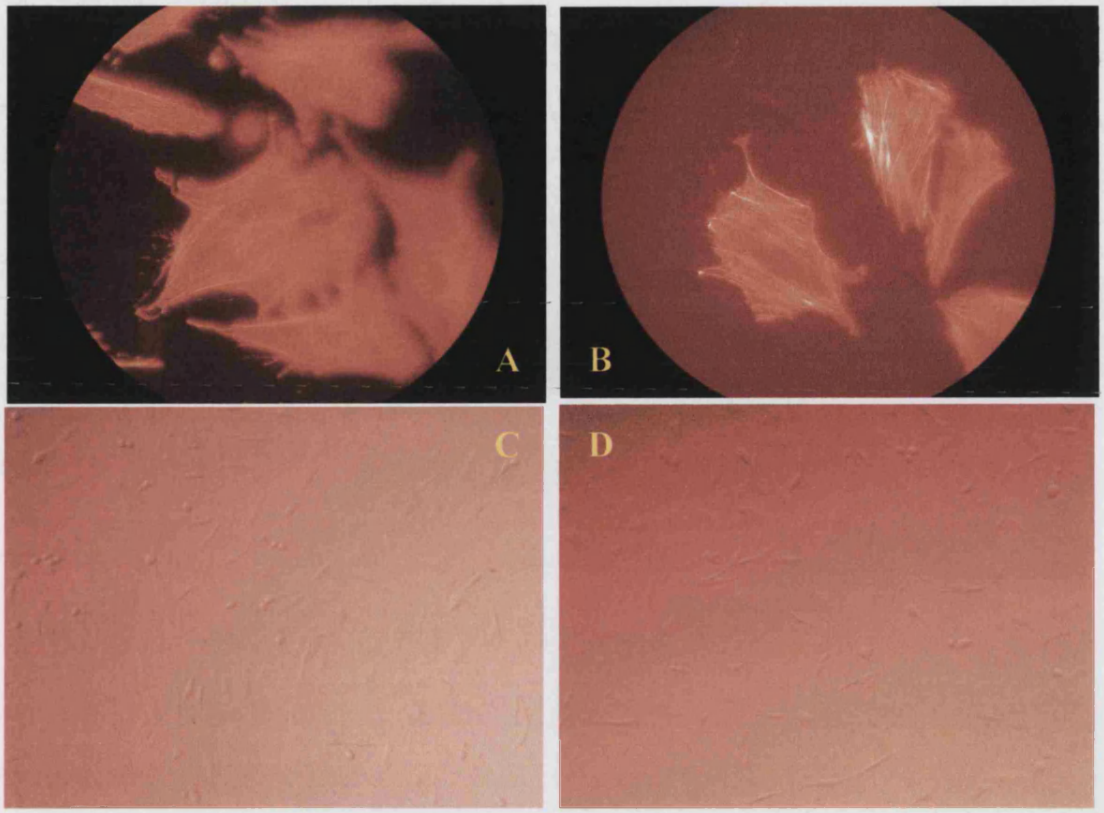


Figure 5.8: NIH 3T3 cell rounding upon addition of C3bot1 and C3stau2. A and C after incubation for 1-3hrs with 15µg/ml C3bot1. B and D after incubation for 1-3hrs with 400µg/ml C3stau2. C and D were visualised on a LEICA DMIRB microscope at 5x magnification. A and B were visualised with an inverted LEICA DMIRB confocal microscope with water emersion at 63x magnification after staining with Rhodamine-phalloidin.

the ADP-ribosyltransferase activity was due to the need to increase the C3stau2 concentration to a level which pushed the activity into the detection range. This high level of C3stau2 meant that it was hard to keep the substrates, particularly RhoA, in excess. Also, the K_D of RhoA for any C3 exoenzyme is unknown. From our observations, it would appear that the similarity in the specific activities of our C3bot1 and C3stau2 must be incorrect as direct comparisons suggest that C3stau2 is less active than C3bot1. This

disagrees with the published results which suggest that both activities are similar. It is possible that the variation in specific activity between our results and the published ones are due to differences in RhoA activity as mentioned previously.

The specific NAD glycohydrolase activity of C3stau2, being approximately 20 times less active than C3bot1, is more consistent with our observed comparison of the two enzymes. It is approximately 100 times less than the published activity (Wilde *et al.* 2002c) for reasons which are unclear. Similarly, the effect of C3stau2 on NIH 3T3 cells was much less than observed previously (Wilde *et al.* 2001). The published results were observed after efficient incorporation using engineered toxins rather than by osmotic shock. It is possible that toxin-mediated uptake optimally localises the C3stau2 near its targets in a way that osmotic shock does not, although this would be assumed to be the same for C3bot1. However, the cell assay results are consistent with the NAD glycohydrolase assays which suggest that C3stau is approximately 100 times less active than the published NAD glycohydrolase activity and our observed ADP-ribosyltransferase activity for C3bot1.

Ignoring the previously published results, the difference in activity between C3stau2 and C3bot1, purified by the same method, could be due to the specificity of these enzymes for RhoA, although this was not quantified by the ADP-ribosylation assays. However, this would not explain the difference in NAD glycohydrolase activities. There are also no obvious clues in the structure to suggest that there is some difference between them. These results may just have to add to the small number of conflicting results on C3 activity which require further investigation.

5.3.4 Investigation of the C3bot1-RhoA complex

Although the structures (Han *et al.* 2001; Menetrey *et al.* 2002) of C3bot1 and mutational analyses of C3bot1 (Menetrey *et al.* 2002) and RhoA (Wilde *et al.* 2000) gave some insight into RhoA binding, the structure of a C3bot1-RhoA complex would answer many of the questions about specificity. This is not a straight forward problem as the interaction is enzymatic and little is known about the binding strength of the complex. Also, mutation of the active site residue of C3bot1, which would prevent ADP-ribosyltransfer, also prevents NAD binding (Saito *et al.* 1995). These factors make it unclear as to whether a complex could be formed that was tight enough for crystallisation trials. The following experiments show the attempts to produce evidence of complex formation in the presence or absence of an NAD analogue.

5.3.4a NAD analogue experiments

The aim of these assays was to screen potential NAD analogues/inhibitors for their ability to promote C3bot1-RhoA binding *in vitro*. Non-hydrolysable NAD can be synthesised but is not routinely available. Therefore, it was hoped that a suitable NAD analogue could be found which would stabilise a C3bot1-analogue-RhoA complex for crystallisation. To assess the usefulness of various NAD analogues to aid stable complex formation, a group of twelve potential inhibitors, were bought or acquired and assessed for their ability to inhibit ADP-ribosylation (Table 5.2). The assay was based on the ADP-ribosylation assay and the ability of the analogue to inhibit the incorporation of labelled ADP-ribose into RhoA.

All assays were done in duplicate, averaged and the results calculated as a percentage loss of C3bot1 activity compared to a positive control as shown in Figure 5.9(A). The analogues showed various degrees of inhibition and the most promising, 1, 5, 6 and 9 were picked for further study. These four analogues were then tested to determine whether they were inhibitors which blocked the active site or were alternative substrates for the enzyme. This was done by incorporating another step into the assay where the samples where pre-incubated for 1hr, before addition of the NAD, either in the presence, or absence of the analogues. If the analogues were alternative substrates for the enzyme there would be a greater apparent loss of activity for the C3bot1. This would be due to the extra time to incorporate the unlabeled analogue into RhoA, reducing the concentration of Rho for modification by the labelled NAD. Again, the results are measured as a percentage of a

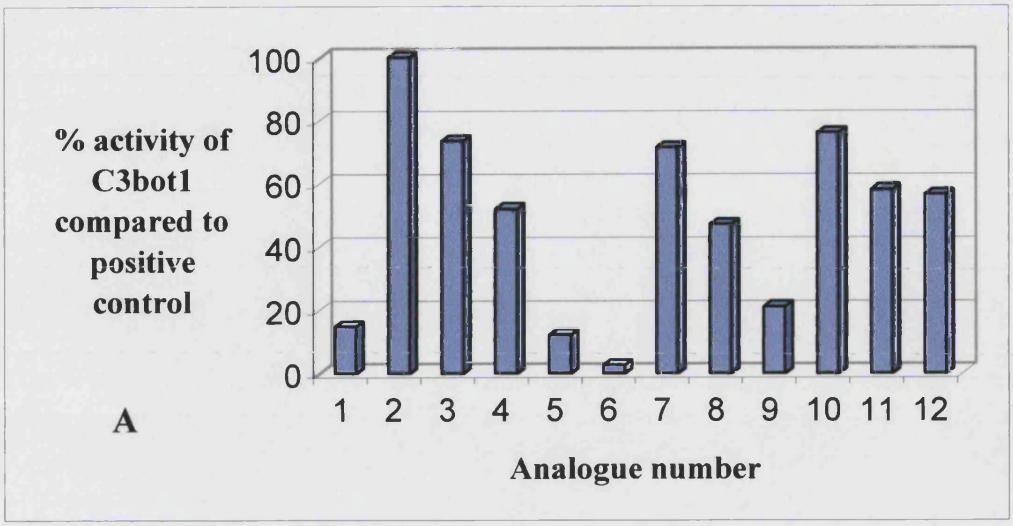
Table 5.2: Analogues screened for inhibition of [adenylate- ^{32} P] ADP-ribose incorporation into RhoA

1.	3-acetylpyridine adenine dinucleotide
2.	3-actelypyridine hypoxanthine dinucleotide
3.	alpha-nicotinamide adenine dinucleotide
4.	nicotinamide 1, N^6 –etheno adenine dinucleotide
5.	nicotinamide guanine dinucleotide
6.	nicotinamide hypothanthine dinucleotide
7.	nicotinic acid adenine dinucleotide
8.	3-pyridinealdehyde adenine dinucleotide
9.	thionicotinamide adenine dinucleotide
10.	adenosine monophosphate (AMP)
11.	β-methylene-thiazole-4-carboxamide adenine dinucleotide (β-TAD)*
12.	selenazole adenine dinucleotide (SAD)*

* A gift from Victor E. Marquez, PhD, Chief, Laboratory of Medicinal Chemistry, Centre for Cancer Research, NCI-Frederick.

positive control as shown in Figure 5.9(B). Unfortunately, the analogues which showed the greatest inhibition, also appear to be alternative substrates for the enzyme as the apparent activity goes down with pre-incubation of the analogue. The most promising looking analogue from this assay, 3-acetylpyridine adenine dinucleotide, required final concentration of 100μM in the reaction indicating a lack of tight binding.

It can be concluded from these experiments that none of the analogues tried are suitable for stabilisation of a C3bot1-RhoA complex as they are either substrates for C3bot1, or do not bind with sufficient strength to promote complex formation.



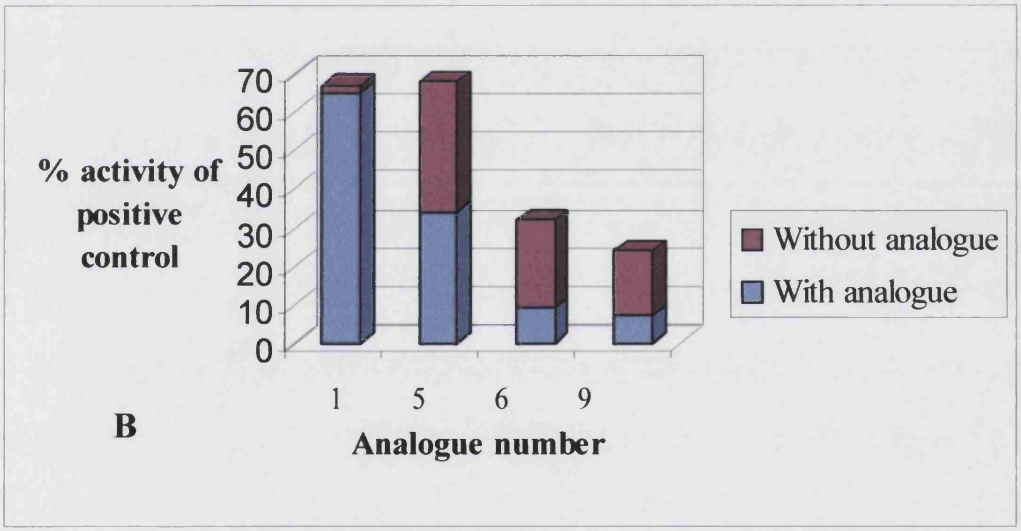


Figure 5.9: A, A bar chart showing the ADP-ribosyltransferase activity of C3bot1 in the presence of the 12 NAD analogues (1mM) shown in Table 5.2 as a percentage of a positive control. **B, A bar chart showing the ADP-ribosyltransferase activity of C3bot1 in the presence of analogues 1, 5, 6 and 9**, pre-incubated for 1hr at 37°C before addition of [*adenylate*- 32P] NAD, either in the presence or absence of the analogue. Concentrations of 1, 5, 6 and 9 were 0.5mM, 0.5mM, 0.2mM and 1mM respectively.

5.3.4b GST binding assays

The characterisation of C3bot1-RhoA binding by Wilde *et al.* suggested an alternative approach (Wilde *et al.* 2000). In this reference, a GST-pull down assay was used to demonstrate C3bot1-RhoA binding in the absence of NAD or an NAD analogue. The assay used glutathione sepharose-bound RhoA-GST to pull down C3bot1 which was detected by western blot. This assay was adapted to use larger quantities of protein than described in the method to allow visualisation of C3bot1 and RhoA by chromatograph or SDS-PAGE. This adapted assay would also more closely mimic the concentrations of proteins needed for crystallisation.

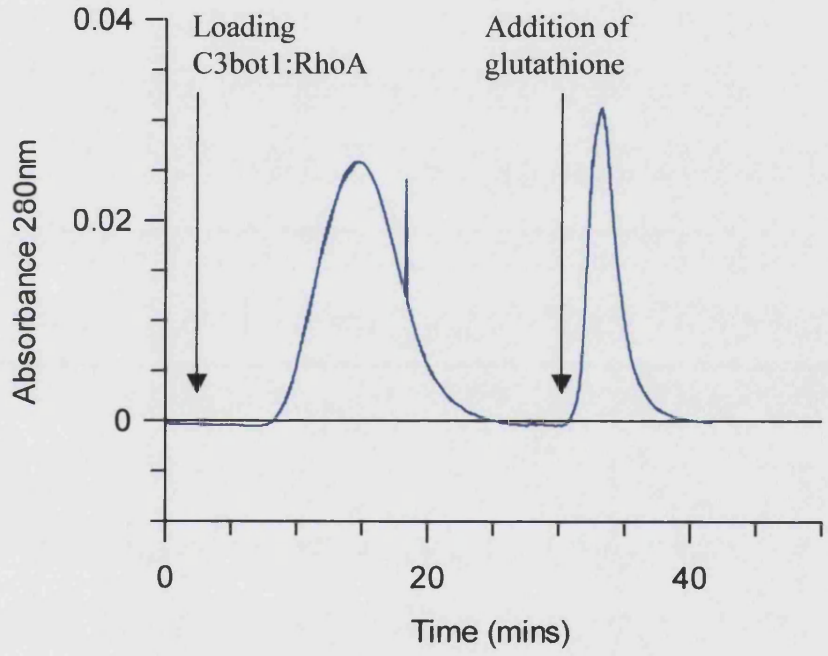


Figure 5.10: Chromatograph for the C3bot1 RhoA-GST binding experiment. C3bot1 and RhoA-GST were incubated overnight at 4°C in 50mM Na.Hepes pH7.3, 2mM $MgCl_2$, 1mM EDTA, 1mM DTT. Bound protein was eluted from the column with glutathione. The peak during loading shows that C3bot1 did not bind.

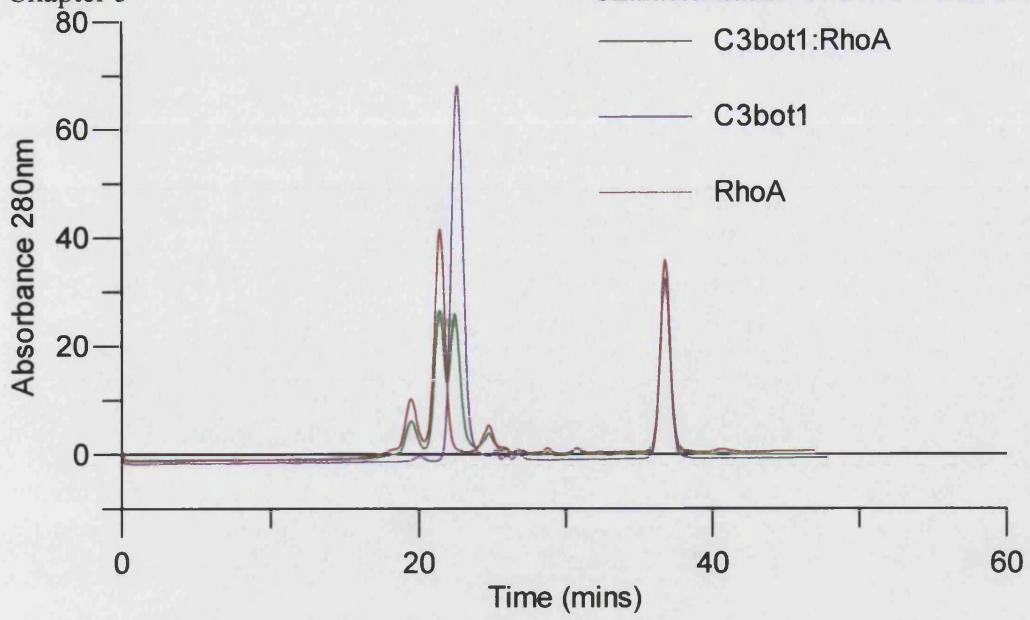
1mg C3bot1 and 0.5mg RhoA-GST in 5mls binding buffer (50mM Na.Hepes pH7.3, 2mM $MgCl_2$, 1mM EDTA, 1mM DTT), were incubated overnight at 4°C. The

mixture was loaded onto a glutathione sepharose column in chilled binding buffer and bound material eluted with binding buffer plus 20mM glutathione. A large peak was observed on the chromatograph during the loading of the protein onto the column (Figure 5.10) suggesting that not all the protein was bound. SDS-PAGE confirmed that C3bot1 localised to this peak and none was observed in the glutathione-eluted RhoA-GST peak. This experiment suggested that any complex formed was not tight enough to withstand the glutathione column chromatography.

5.3.4c Gel filtration binding assays

Subsequent to the failure of the GST-capture assay, gel filtration was an obvious choice for purification of a complex. Initial experiments were carried out with a Superdex 200 HR 10/30 gel filtration column using the same buffer as used for the GST-capture assays (50mM Na.Hepes pH7.3, 2mM $MgCl_2$, 1mM EDTA, 1mM DTT) with the addition of 0.1M NaCl to prevent binding of the proteins to the column. Varying concentrations of the proteins were used to monitor any difference in complex saturation depending on which protein was in excess. The results of these experiments, although confused by impurities in the RhoA causing peaks in the chromatograph, clearly showed no evidence of a C3bot1-RhoA complex (Figure 5.11) with excess C3bot1 or RhoA.

As a final attempt to purify a C3bot1-RhoA complex, the gel filtration assay was repeated in the presence of 150μM 3-acetylpyridine adenine dinucleotide. Although this concentration is not as high as required for the inhibition assay, it was the highest that could be achieved using a minimal 50mls of running buffer. It was hoped that this analogue may show some ability to produce a transient complex as even transient complexes can be measured by the change of peak positions on the chromatograph (Beeckmans 1999). A Zorbax Bio Series GF-250 column was used for this assay to reduce the quantity of the sample and the volume of running buffer required. Although these results were also slightly confused by an unclear trace for C3bot1, possibly due to its adhesion to the column at only 0.1M NaCl, it is clear that complex formation was not promoted by the 3-acetylpyridine adenine dinucleotide (Figure 5.12). The lack of success in these assays suggests that if C3bot1 and RhoA can form a complex, *in vitro*, in the absence of NAD, then it is not a particularly tight complex. It is not tight enough to detect by glutathione pull-down in the absence of salt, or by gel filtration in the presence of salt at concentrations below physiological levels.



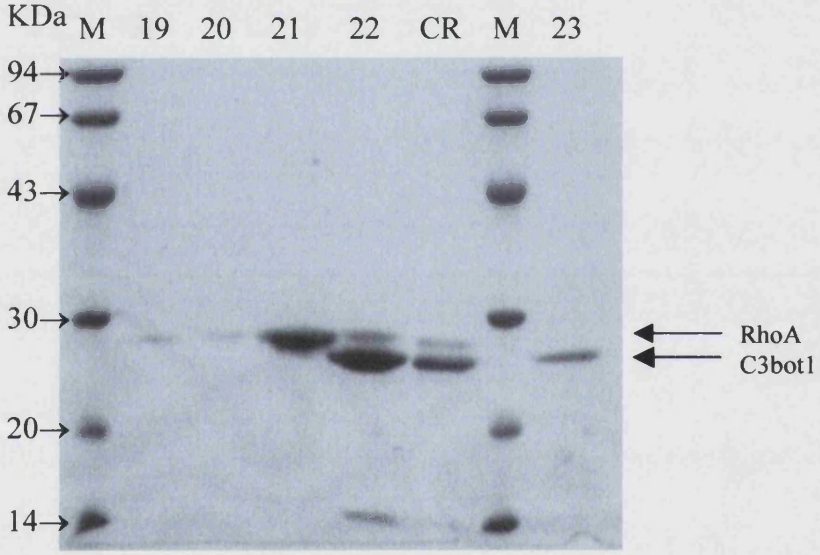


Figure 5.11: A, Overlay of the chromatographs from the C3bot1 RhoA binding experiment. Individual samples of C3bot1 and RhoA as well as a mixture incubated overnight at 4°C in 50mM Na.Hepes pH7.3, 2mM $MgCl_2$, 1mM EDTA, 1mM DTT, were separated by gel filtration with a Superdex 200 HR 10/30 column (Amersham). No peaks were observed in the mixed sample which did not correspond to C3bot1 or RhoA. The peak on the right is a buffer effect. **B, 12% SDS-PAGE gel of the C3bot1 RhoA binding experiment.** M is the low molecular weight marker (Amersham), CR, a sample of C3bot1 and RhoA and the numbers relate to the fractions. RhoA is found in the first peak corresponding to fractions 19-22 and C3bot1 in the second, corresponding to fractions 22-23.

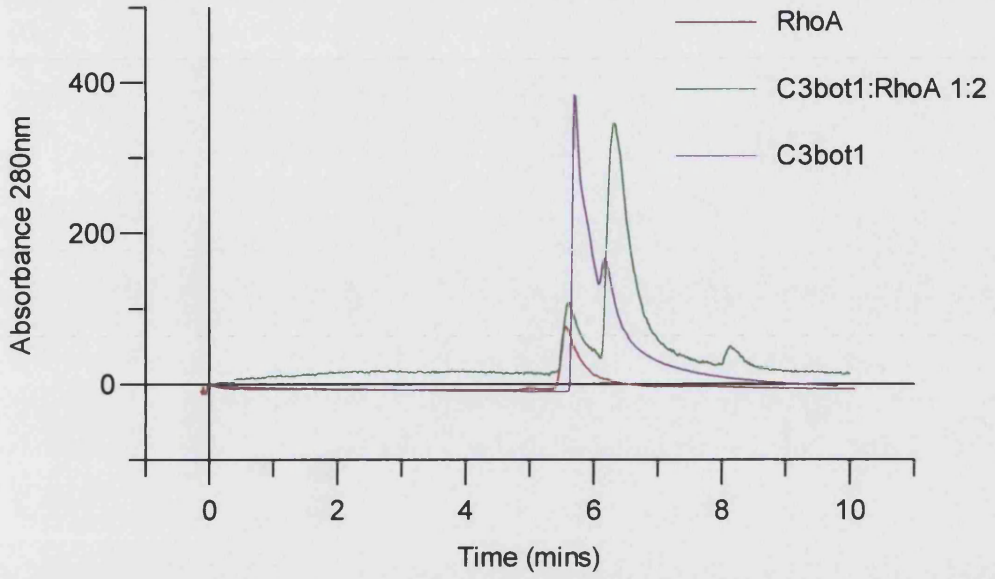


Figure 5.12: Overlay of the chromatographs for the C3bot1 RhoA binding experiment. Individual samples of C3bot1 and RhoA, as well as a mixture incubated overnight at 4°C in a molar ratio of 1:2, were passed over the Zorbax Bio Series GF-250 gel filtration column in 50mM Na.Hepes pH7.3, 0.1M NaCl, 5mM $MgCl_2$ and 1mM DTT with 150μM 3-acetylpyridine adenine dinucleotide. The mixed sample shows no peaks preceding, and therefore corresponding to a molecular weight larger, than C3bot1 or RhoA, suggesting that a complex was not formed.

5.3.4d Surface plasmon resonance experiments

As chromatographic experiments had not identified a C3bot1-RhoA complex, the nature of the binding was investigated further by surface plasmon resonance. The affinity of protein complexes can be measured by the direct interaction of two proteins monitored using surface plasmon resonance (SPR) through a Biacore machine. The surface plasmon resonance of a protein 'ligand' chip changes on the formation and dissolution of complexes when a liquid 'analyte' sample is passed over. This allows kinetic measurements of low affinity binding interactions and the affinity of second substrates in two substrate reactions to be measured. The high sensitivity of this technique makes it ideal for measuring transient enzymatic complexes such as C3bot1-RhoA.

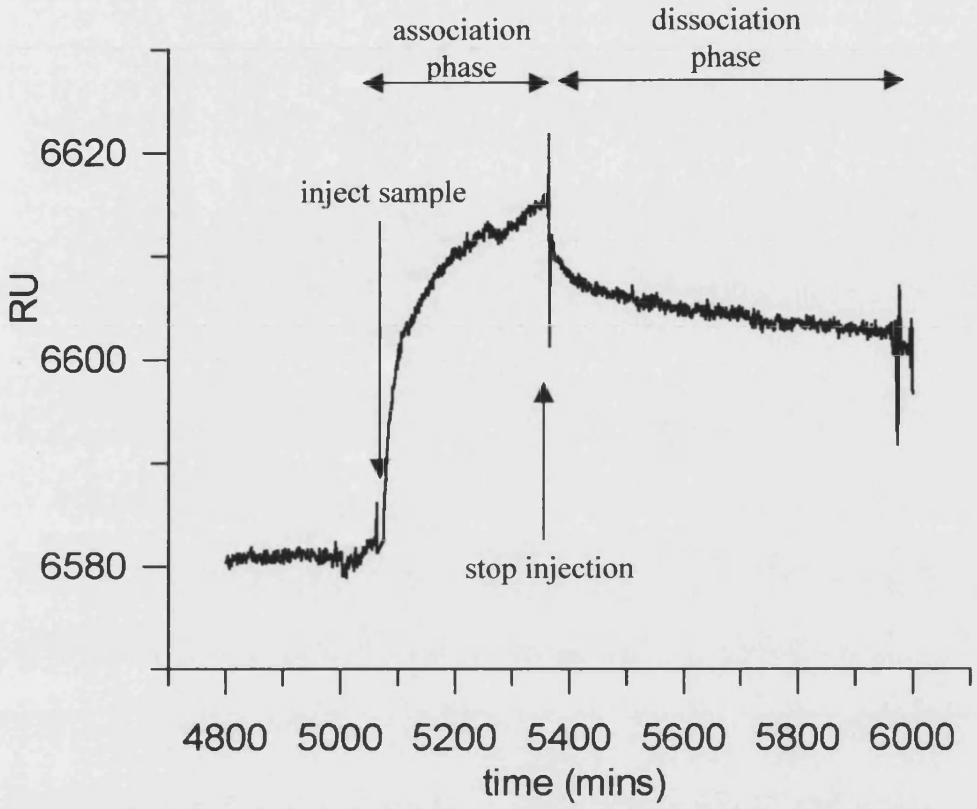


Figure 5.13: A typical Biacore trace. The surface plasmon resonance response units increase during sample injection as the analyte binds the ligand. After injection is finished, the ligand equilibrates with the buffer passing over the chip and the response units decrease.

A C3bot1 chip was successfully made with amine-coupling resulting in approximately 6.3μg protein bound to the chip as detected by a change in the running baseline of 6,256 SPR units. The kinetics of binding are measured by the association and dissociation of the Biacore trace (Figure 5.13) but the total mass of analyte bound corresponds to the increase in SPR units before dissociation. Amine coupling does not guarantee a favourable orientation for the C3bot1 on the chip but the binding of a similar size protein (such as RhoA) would be expected to give and increase in response units of up to 50% of the difference seen for C3bot1 on chip creation. Injection of 100μl 1mg/ml

RhoA over the chip at 50µl/min, however, only resulted in a response difference of 20 units (0.3% of the C3bot1 response of 6000). This, along with some RalA experiments suggested that only a very small fraction of the C3bot1 on the chip was able to bind analytes or that the affinity for RhoA in the absence of NAD was very low. Although further experiments could be carried out with NAD, it would be hard to optimise binding of two substrates in the absence of a positive control to show the maximum binding.

These results lead to the design of a second experiment using a RhoA-GST chip in the hope that the amine coupling would leave more of the RhoA free to bind the C3bot1. Unfortunately the largest response difference observed for injection of 390µg/ml C3bot1 at 50µl/min was ~100 units (<1% of activated surface of 12,000) which did not seem to be large enough to indicate either an active chip or tight binding as injection of BSA produced a similar response.

Although Biacore would seem to be the ideal way to study these reactions, the time and expense required to get a functional assay system was beyond the scope of the size of this PhD project. It is possible to use carefully added tags to position proteins optimally on the chip surface but this requires much work and often trial and error using several (expensive) chips. It is also necessary to have good controls to analyse the total active protein on the chip and to reduce noise from non-specific binding. Further experiments would have to consider these factors carefully.

5.3.4e Discussion

Although these experiments were performed with a variety of different RhoA samples, the total lack of success of any of these experiments suggests that the binding strength of any C3bot1-RhoA interaction in the absence of NAD is likely to be low and would not lend itself to crystallisation. It is also unfortunate that none of the NAD analogues tried appeared to promote this interaction and mainly acted as substrates. The best option to stabilise a C3bot1-RhoA complex would be with a non-hydrolysable NAD analogue such as carbanicotinamide adenine dinucleotide (Slama and Simmons 1988). However, either this experiment or further Biacore experiments would prove to be expensive.

5.3.5 Investigation of the C3bot1-RalA complex

The report that the binding affinity of the C3lim-RalA complex was 12nM (Wilde *et al.* 2002a), coupled with the evidence that RalA inhibited C3bot1 and C3lim functions to similar levels, suggested that the C3bot1-RalA complex would be suitable for characterisation with our techniques. The following experiments show the attempt to characterise the C3bot1-RalA complex for crystallisation.

5.3.5a Gel filtration binding assays

To attempt to purify a C3bot1-RalA complex for crystallisation, the previously described gel filtration binding assays were repeated with RalA. The binding assays were performed with the Superdex 200 HR 10/30 column in the published binding buffer (50mM Na.Hepes pH7.3, 2mM $MgCl_2$, 1mM DTT, 0.5% NP40) plus 0.1M NaCl to prevent protein adhesion to the column. 2.6nmol RalA were incubated with either 5.1nmol or 1.3nmol C3bot1 in a volume of 100µl buffer overnight at 4°C before separation by gel filtration. Column fractions were concentrated by precipitation with TCA before separation by SDS-PAGE. Figure 5.14 shows the chromatograph and gel for the 1:2, C3bot1:RalA ratio. Although it is difficult to spot the peaks above the noise on the chromatograph due to the small amounts of protein used, the gel makes it clear that none of the fractions contain C3bot1 and RalA indicating that a C3bot1-RhoA complex was not formed. This was also the case for the reverse experiment.

The lack of complex formation suggests that either that the additional salt added for the gel filtration experiment interferes with the complex, that the affinity of RalA for C3bot1 is not as high as presumed, or that the RalA was denatured in some way. The first of these would seem to be the least likely, so the obvious choice was to test RalA binding to C3bot1 using Biacore.

5.3.5b Surface plasmon resonance experiments

The binding of RalA to C3bot1 was analysed by Biacore using the C3bot1 chip and the running buffer described previously for Biacore experiments. Unfortunately, injection of 100µl 0.15mg/ml RalA-GST at 50µl/min only resulted in a response difference of 35 units suggesting that either this chip or the RalA were only minimally active or that the binding strength of this interaction is not above background.

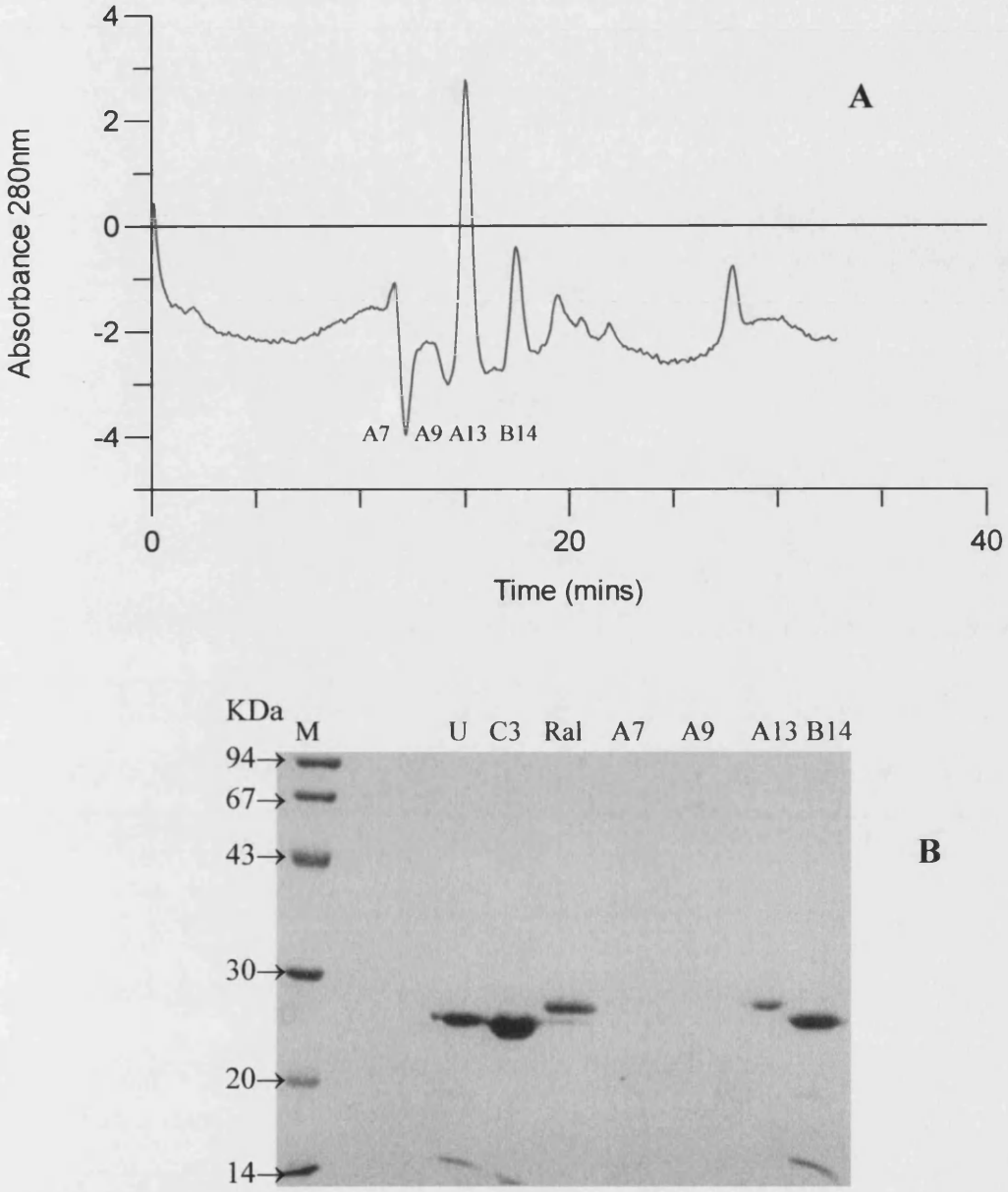


Figure 5.14: A, Chromatograph of the C3bot1 RalA binding experiment. C3bot1 and RalA were incubated overnight at 4°C in 50mM Na.Hepes pH7.3, 2mM $MgCl_2$, 1mM DTT, 0.5% NP40. **B, 12% SDS-PAGE gel of C3bot1-RalA binding.** M is the low molecular weight marker (Amersham), U, a sample from another experiment, C3 and Ral, C3bot1 and RalA markers and A7-B14 their respective fractions. Peaks A7 and A9 are noise with no protein associated with them. RalA and C3bot1 are found in fractions A13 and B14 respectively. No C3bot1-RalA complex is observed.

5.3.5c Discussion

Although the reported strength of the C3bot1-RalA interaction is in the nM range, no complex formation was observed by gel filtration or Biacore assays. The Biacore assays cannot be held to be conclusive as they were not optimised and suitable controls were not found. The reason for the failure of the gel filtration assay is less certain unless the binding strength of the interaction is less than expected, or the RalA is inactive. Although the GTPase activity of the RalA was not tested, the purification was identical to that of RhoA, which was active. However, similar proteins do not always have similar activities as shown by the differences between C3bot1 and C3stau2. More time to optimise the purification and activity of this enzyme would be needed for further investigation of the binding specificity.

Chapter 6

Crystallisation of Botulinum neurotoxins H_C/F and H_C/A

6.1 Introduction

Botulinum toxins are now widely used therapeutically after the success of BoNT/A in correcting strabismus (Johnson 1999) due to its the effectiveness in weakening muscle. BoNT/A is licensed in the United Kingdom for treating blepharospasm, hemifacial spasms, cervical dystonia and axillary hyperhidrosis (Misra 2002) although there are many reports of success with other conditions. BoNT/B is also licensed in the U.S.A for cervical dystonia (Klein 2003). Botulinum toxins are ideal nerve blocking agents due to their specificity for cholinergic neurones, the reversibility of their action and because their mechanism is understood (Johnson 1999). Successful injection into the required area can be achieved without undue side effects on surrounding neurones. However, their main draw back is an induction of an immune response with repeated injection. Although the effect of BoNT/A can last up to three months (Munchau and Bhatia 2000), other neurotoxins such as BoNT/B, E, F and C, which can be as effective as BoNT/A, do not have such long lasting effects (Eleopra *et al.* 1998; Houser *et al.* 1998; Foran *et al.* 2003).

The distinct domain structure of the BoNTs and the high level of specificity of both the heavy chain and the light chain has also lead to the possibility of using individual domains of the toxin as chimeric therapeutics. For example, the light chain and translocation domain fragment of BoNT/A, LH_N/A, has been engineered with the *Ethyrina crystagalli* lectin to change its cell binding specificity to disable nociceptive neurones for the prevention of pain (Duggan *et al.* 2002). The other possibility is to use the binding domains to target other proteins to neuromuscular junctions. Most research in this field has been done with the tetanus H_C domain, which has been shown to be capable of retrogradely transporting chimeric proteins to the motor neurones (Coen *et al.* 1997) and coupled with diphtheria toxin can be effectively transported into the cytosol (Francis *et al.* 2000). The tetanus H_C domain has been engineered to carry various therapeutics including β-*N*-acetylhexosaminidase A for the prevention of lysosomal storage diseases (Dobrenis *et al.* 1992), the neuroprotective superoxide dismutase (Figueiredo *et al.*) and adenoviral vectors (Schneider *et al.* 2000). Recently, it has been coupled with diphtheria toxin and a DNA-binding domain for translocation to the nucleus (Barati *et al.* 2002).

Much less work has been carried out on BoNT fragments and only one construct has been reported using non-recombinant, cleaved, toxin (Goodnough *et al.* 2002). The

reason for this contrast in research is that, unlike tetanus toxin, it is difficult to express active recombinant BoNT H_C fragments. Of the BoNTs, only recombinant H_C/F and H_C/C showed any neuronal binding and that, much less than the native toxin (Sutton, J.M., Shone, C.C., -unpublished results). The reason for this disparity in unknown. The exact nature of the ganglioside binding site for the BoNTs is also not fully understood and the this site has mainly been studied in tetanus toxin where the H_C domain has been crystallised with various sugars and a ganglioside GT1b analogue (Emsley *et al.* 2000; Fontinou *et al.* 2001).

Therefore, in order to understand the difference in affinity of H_C/F domain for gangliosides, in the absence of the rest of the toxin, we attempted to crystallise the H_C domain of BoNT/F, and also BoNT/A for comparison. These experiments were also hoped to increase our understanding of the ganglioside binding site for these toxins. This chapter describes the unsuccessful crystallisation attempts for H_C/F and H_C/A and comments on the solubility of the proteins for use in further trials.

6.2 Methods

BoNT/F and BoNT/A H_C – These were purified as described previously (Shone and Tranter 1995) and supplied in 10mM Hepes, pH 7.2, 0.2M NaCl, 2mM DTT at a concentration of 5mg/ml, by Dr. C.C. Shone at the HPA.

Hanging drop method- 1.5µl well solution was mixed with 1.5µl protein and suspended above 0.7/1ml reservoir. The drops were incubated at 4°C, 16°C or 22°C.

Hanging drop method with oil- As above, but 500µl of 50:50 parafin:silicone oil was layered over the well solution to slow the diffusion between the drop and the precipitant.

Crystallisation under oil- 1µl precipitant was mixed with 1ul protein and eluted under oil in a microbatch plate. The drops were incubated at 16°C.

6.3 Results

Initial screening was carried out for both BoNT/F H_C and BoNT/A H_C with the commercially available screens at the maximum concentration of protein (5mg/ml) and incubated at 16°C. These screens represent a range of a hundred conditions which have commonly given crystals in previous experiments. The commercial screens produced many drops with heavy amorphous precipitate for both proteins suggesting that either the protein or the precipitant concentration was too high in most cases. Table 6.1 shows the only precipitant from the 100 screened that did not precipitate the proteins. For H_C/F, only 10 out of a 100 precipitants gave clear drops. For H_C/A, this number was not much better with only 17/100. To get a better idea of the concentration of a variety of precipitants required to precipitate the protein, the commercial screens were carried out at half the precipitant concentration. These screens proved to be at a more suitable concentration for H_C/F as just under half the precipitants did not precipitate the protein and the ones that did gave a range of precipitation levels. Although for H_C/A, only a quarter were clear drops, most drops only had small amounts of precipitate, some of which were more crystalline in nature. These screens were also set up as microbatch experiments to try out this method and to see if the lack of vapour diffusion gave crystals for these conditions. In general, these screens gave very similar results to the vapour diffusion experiments. Also, they were more time consuming to set up so this approach was abandoned. None of the conditions tried produced crystals of H_C/A, but a small crystal was observed in a H_C/F screen with 0.4M Na/K Tartrate after 3 months. Unfortunately, this crystal dissolved when extra mother liquor was added to the drop. Repeated experiments of this and similar conditions produced no crystals.

Following the lack of success with the commercial screens to produce any preliminary crystals, screens based on the published conditions for the crystals of BoNT/A and BoNT/B were tried. Screens based around the conditions of the tetanus toxin H_C crystallisation were also extensively tried. Tetanus toxin H_C, in 20mM Imidazole, pH 7, 100mM NaCl was crystallised at 2.8-3mg/ml in 200mM ammonium sulphate with 40% PEG 4000 and 1% MPD (Emsley *et al.* 2000), 20% PEG 4000, 0.2M Imidazole, pH 7 (Fontinou *et al.* 2001), or 20% PEG 8000, 0.2M Na_2KPO_4, 0.1M Tris, pH 8.5 (Fontinou *et al.* 2001). However, at 3mg/ml, both F and A H_C were very strongly precipitated by these conditions at 16°C. Variations on these conditions did not produce crystals.

BoNT H_C/F	BoNT H_C/A
0.4M Na/K Tartrate	1M Imidazole
1.0M Ammonium phosphate 0.1M Na Citrate pH 5.6	20% 2-propanol, 0.1M Hepes pH 7.5 0.2M Na Citrate
0.1M Na Acetate,0.1M Imidazole pH 6.5	30% PEG 1500
3.4M 1.6, Hexanediol 0.1M Tris pH 8.5, 0.2M Mg Chloride	30% MPD, 0.1M Na Acetate pH 4.6 0.02M Ca Chloride
4.3M Na Chloride, 0.1M Hepes pH 7.5	2M Urea
2M Ammonium Formate 0.1M Hepes pH 7.5	1M Ammonium dihydrogen phosphate 0.1M Na Citrate pH 5.6
2M Na Chloride, 0.1M Mes pH 6.5 0.2M Na/K phosphate	18% PEG 8000, 0.2M Ca Acetate 0.1M Na Cacodylate pH 6.5
0.5M Na Chloride, 0.1M Mg Chloride 0.01M Cetyl trimethyl Ammonium Bromide	3.4M 1,6, Hexanediol, 0.1M Tris pH 8.5 0.2M Mg Chloride
4% polyethyleneamine 0.1M Na Citrate pH 5.6 0.5M Na Chloride	10% Jeffamine M-600 0.1M Na Citrate pH 5.6 0.01M Ferric Chloride
10% Jeffamine M-600 0.1M Na Citrate pH 5.6 0.01M Ferric Chloride	4% polyethyleneamine 0.1M Na Citrate pH 5.6 0.5M Na Chloride
	30% Jeffamine M-600, 0.1M Mes pH 6.5 0.05M Caesium Chloride
	2M Na Formate, 0.1M Hepes pH 7.5
	2M Mg Chloride, 0.1M Bicine pH 9
	0.1M 1,6, Hexanediol, Na Acetate pH 4.6 0.01M Co Chloride
	2M Ammonium Sulphate, Tris pH 8.5
	10% Ethanol, 1.5M Na Chloride
	8% PEG 4000, 0.1M Na Acetate pH 4.6

Table 6.1: Precipitants from the initial commercial screens which did not precipitate the proteins at 16°C with a protein concentration of 5mg/ml. Some metals are written using their chemical symbol e.g. Na, sodium, Mg, magnesium, K, potassium, Ca, calcium, Co, cobalt.

Crystal nucleation occurs at the boundary between the unsaturated protein solution and the supersaturated solution which precipitates the protein (see Chapter 1). Conditions which only precipitate the protein slightly, are probably not far, 'in crystallisation space', from those that might crystallise the protein. Therefore, the next strategy, was to take conditions which only precipitated the proteins slightly, and to vary and manipulate them with the addition of additives. Additives included glycerol, sucrose, ethanol, ethylene glycol, zinc acetate and sodium orthovanadate, as well as small amounts of detergents such as octyl-β-D-glucopyranoside (βOG), Triton X100 and CHAPS. These experiments produced many clear and precipitated drops. Detergents have been reported to be useful at low concentrations as additives (McPherson *et al.* 1986a; McPherson *et al.* 1986b) rather than the higher concentrations required for solubilising membrane proteins. In particular, βOG has been heralded as a particular success (McPherson *et al.* 1986a; McPherson *et al.* 1986b). Of the detergents tried, Triton X100 was the only one which produced a light crystalline precipitate and not amorphous precipitate. βOG usually resulted in large amounts of amorphous precipitate. However, none of these trials produced crystals. For some of the H_C/A conditions that had the most promising crystalline precipitates, trials with a layer of 50:50 parafin:silicone oil was layered over the well solution in an attempt to slow vapour diffusion and therefore give the protein more time to crystallise. However, this technique only varied the amount of precipitate and did not produce crystals.

The final strategy was to find more 'borderline' conditions by screening precipitant concentrations between those that had precipitated the protein in the first set of commercial screens but not in the second at the lower concentration. These were mainly PEG/salt against pH grid screens set up either at 16°C or 22°C. Table 6.2 shows the boundary conditions found from these screens at 22°C. H_C/F was more widely screened that H_C/A due to the availability of protein. Some of these precipitants were screened with quite finely spaced conditions but none produced crystals.

From the various screens there are some general points to be noticed. Both proteins was much less soluble at pH 5.5 and this is presumed to be due to a change in conformation, exposing hydrophobic regions, for membrane insertion. Also, H_C/F particularly lacks solubility in alcohols and is also less soluble in PEG than salts at the

concentrations usually used for crystallisations (e.g. in the commercial screens). H_CA produced many more crystalline looking precipitates than H_C/F suggesting that it may be more amenable to crystallisation, although without crystals this is only speculation.

BoNT/F H_C	**BoNT/A H_C**
~18% MPD, 0.1M Na Acetate, pH 4.6	~ 1.2M Ammonium sulphate, 0.1M Tris, pH 8.5
~1.2M Ammonium sulphate 0.1M Tris, pH 8.5	~ 3% PEG 8000, 0.1M Tris, pH 8.5
~0.8M Na Acetate, 0.1M Na Cacodylate pH 6.5	
12-18% PEG 400, 0.1M Hepes, pH 7.5	
0.8-1M Li Sulphate, 0.1M Imidazole/malate, pH 7	
~ 2.3M Na Formate, 0.1M Imidazole/malate, pH 7	
~ 0.8M Na Citrate, 0.1M Imidazole/malate, pH 7	
~3% PEG 8000, 0.1M Imidazole/malate, pH 7	
~6% PEG 8000, 0.1M Na Acetate, pH 4.5	

Table 6.2: Boundary conditions or ranges between no precipitation and slight precipitation for 4mg/ml H_C/F and H_C/A at 22°C. Sodium and lithium are abbreviated to their chemical symbols Na and Li, respectively.

6.4 Discussion

Crystallisation of proteins requires finding the right set of conditions to supersaturate the protein solution in the most favourable way for crystallisation. Of the factors that effect the solubility of the protein, and therefore its saturation level, protein concentration, precipitant concentration and type, pH and temperature are the easiest to vary in crystallisation experiments. Of these, however, the hardest to screen is temperature as incubators are usually available at set temperatures so screening is usually carried out at just one temperature. In these experiments crystallisations were mostly set up at 16°C or 22°C. Although temperature is critical in crystallisation experiments, if it is held constant, the other factors can be manipulated to supersaturate the protein at that temperature. For ease of screening, the protein concentration is also often held constant and the precipitants and pH are varied. However, varying the protein concentration slightly was tried for some experiments.

In designing crystallisation trials, it is often hoped that the preliminary commercial screens will produce some form of preliminary crystals that can be optimised by adjusting the aforementioned parameters. In the absence of preliminary crystals, further screens must be designed to sample the 'crystallisation space' and to find the solubility boundaries at which crystals may grow. Without experience in complex statistical analyses, this task is often hard to do effectively by one person with a pencil and paper. Another way to explore the crystallisation space more effectively is to set up more screens. The use of robots for setting up large trials, and the improvements in computerised crystal identification techniques to observe these trials can also improve the chances of finding suitable conditions. In the absence of a robot, experiments must be designed as well as possible to sample the crystallisation space whilst conserving protein and time required to set up the experiments. The strategies used in these experiments became more effective throughout the trials as experience was gained designing trials and scoring drops for their precipitate quality. However, there is still a large proportion of the crystallisation space left unexplored. The lack of obvious leads meant that time was spent studying small parts of the space without the guarantee of crystals as an end product. Although these experiments produced no crystals, they have gone some way towards describing the solubility of these proteins for future screens.

Conclusions and further work

C3 exoenzyme Research

The published structures of C3bot1 (Han *et al.* 2001; Menetrey *et al.* 2002) and C3stau2 (Evans *et al.* 2003) have identified key residues in NAD and RhoA binding and helped to classify the C3 and binary toxins as mechanistically distinct from the DT and CT-like toxins. However, there is still some confusion about the enzymatic mechanism of these proteins. Comparisons of active site features has been a useful tool in analysing the ADP-ribosylating toxins and it has often been assumed that they share a common mechanism for catalysis. The most detailed work to date on the transition state complex of an ADPRT is the characterisation of the Pertussis toxin active site using kinetic isotope effects. This research suggested an S_N2-like mechanism involving a direct attack on the NC1 carbon atom (Figure 2.1, Chapter 2) concurrent with cleavage of the glycosidic bond (Scheuring and Schramm 1997b; Scheuring and Schramm 1997a; Scheuring *et al.* 1998). The structures of VIP2 (Han *et al.* 1999) and C3bot1 (Han *et al.* 2001) were therefore analysed with the assumption of a similar mechanism. The authors of the crystal structure of Iota toxin, however, are convinced that their structural and mutational analysis is more consistent with an S_N1 mechanism (Sakurai *et al.* 2003; Tsuge *et al.* 2003). In this theory, the positioning of the NAD in a folded conformation allows the phosphate NO1 atom to temporarily hydrogen bond to the nicotinamide NN7 atom resulting in spontaneous cleavage of the glycosidic bond. The resulting oxocarbenium cation is then stabilised by the catalytic glutamate and the first serine from the STS motif until transfer to the protein substrate. So far there has been no direct evidence for either of these theories for either the binary toxins or the C3-like exoenzymes.

It is possible that the mechanism of NAD hydrolysis varies between the two groups as the first serine of the STS motif has been shown to be essential for hydrolysis in the binary toxins (Barth *et al.* 1998; Nagahama *et al.* 2000) but not in the C3-exoenzymes (Menetrey *et al.* 2002; Wilde *et al.* 2002b). This is particularly worth considering as it has been shown that equivalent residues in DT and CT are not interchangeable suggesting different mechanisms of NAD hydrolysis between these two groups (Dolan *et al.* 2000). Similar detailed study of these enzymes, such as using kinetic isotope exchange may be needed to resolve these questions. Alternatively, the structure and kinetic analysis of a C3

exoenzyme or binary toxin with a transition state analogue or other substrate analogues may help illustrate some of these points.

The reason for the difference in catalytic rates of these enzymes is unclear. Even putting aside experimental differences, there are reported comparative differences between the C3 exoenzymes (Wilde *et al.* 2002b) as well as our own observed differences. More detailed kinetic analysis may help answer some of these questions. As far as our C3stau2 is concerned, an ADP-ribosylation assay with a longer incubation time which could use less enzyme would probably be the best way to get an accurate measurement of the activity. Alternatively, a time course, such as that used for the NAD glycohydrolase activity may be helpful. Additionally, an independent assessment of RhoA activity with the assay may be useful. The difference in C3 affinity *in vitro* for RhoB reported (Wilde *et al.* 2002b) is also unexplained. The valine to isoleucine swap at position 43 has been shown to be the reason for the decrease in ribosylation of RhoC. RhoB only differs from RhoA in the first 80 residues at position 10 where it has a valine for isoleucine swap and position 29 where is has a glutamate rather than a glutamine. The role of these residues in the reaction could be tested by mutational analysis.

The specificity of C3 exoenzymes for Rho GTPases is still not understood. A crystallographic C3-RhoA complex would prove hard to achieve without a tight binding non-hydrolysable NAD analogue. An alternative might be to make an ADP-ribosylated RhoA-C3 complex at pH5 in the absence of nicotinamide. Some thought would need to be put into purifying 100% ribosylated RhoA, however. In the absence of these, modelling studies of the NAD-bound form of C3bot1 or C3stau2 with the GDP bound form of RhoA may give some clues. However, as both these molecules have proven to be flexible it is possible that either one, or both, of them may change conformation again on binding. If C3 cleaves the NAD before RhoA binding as suggested by the S_N1 mechanism, this complex is likely to be in a conformation different again to that seen in the crystal structures for docking to Rho. The structure of C3stau2 has given valuable insight into the difference in specificity with C3bot1. However, further mutational analysis based on this structure may help identify further residues in Rho recognition. As EDIN A has now been shown to share the same specificity as C3stau2 (Wilde *et al.* 2003), a comparison of these sequences, combined with the structure of C3stau2 may also help discriminate residues which are likely to be responsible for these differences.

Biacore experiments may also useful for determining the affinity of the C3s for RhoA or RhoE if suitable chips could be made. Direct comparisons would probably be more difficult, however, as a RhoA chip could not be used in conjunction with NAD. It is possible that a uniformly ribosylated RhoA chip investigated at pH5 may give some insight. Additionally, the role of C3-Rho binding in the absence of NAD is not only unclear but also not well characterised. Questions remain as to how far this interaction resembles the complex with NAD and involves the likely conformational changes. This phenomenon needs to be more widely studied to assess its usefulness in understanding the ribosylation reaction. Biacore and NMR experiments would be ideal for studying this interaction.

There is plenty of scope for further investigation of C3-RalA interactions. This interaction should be easier to study due to the reported higher affinity of this complex (Wilde *et al.* 2002a). A crystal structure of a C3bot1-RalA complex could also be used to investigate the specificity of this reaction with by subsequent modelling of C3stau2 with RalA. Biacore studies of both these complexes would give valuable information about these interactions if a suitable chip could be made. However, further research into a role for inhibition of RalA in cell disruption is essential. At present, there is a lack of evidence for such a role and the tight nature of the interaction would suggest that it is more likely to act as an inhibitor of C3, mopping it up in the cell.

Finally, the role for C3 exoenzymes in pathogenesis still needs elucidating. So far, the only suggestions for possible cell entry mechanisms are for the EDIN (C3stau) subfamily and none have these have been tested. The role of Rho in regulating immune functions have lead to the proposal that C3-like exoenzymes may have a possible role in targeting the immune system (Wilde and Aktories 2001) but this hypothesis has yet to be addressed directly.

Binding affinity and interactions of H_C/F and H_C/A and chimeric proteins

The lack of success of this project means that answers to questions of recombinant and chimeric binding domain affinities remain unanswered. Since the start of this project,

the affinity of recombinant BoNT/C H_C for gangliosides has been shown to be much closer to the wild type than the other recombinant H_Cs. However, there is still plenty of scope for further research into BoNT H_C specificity and binding affinity as most of the work in these areas has been done for tetanus toxin.

For the crystallisation of the H_C fragments, these experiments did not produce many leads as to where to go next. As the precipitants tried have not produced crystals, a more broad screening of other precipitants, using suitable concentrations based on the known solubility, may be the best next step to find more leads. Subsequently, streak seeding of the more crystalline precipitates into slightly lower precipitant concentrations might be tried to enhance crystal growth.

References

Adam-Vizi, V., S. Rosener, K. Aktories and D. E. Knight (1988). "Botulinum toxin-induced ADP-ribosylation and inhibition of exocytosis are unrelated events." FEBS Lett. **238**: 277-280.

Aepfelbacher, M., M. Essler, E. Huber, M. Sugai and P. C. Weber (1997). "Bacterial toxins block endothelial wound repair. Evidence that Rho GTPases control cytoskeletal rearrangements in migrating endothelial cells." Arterioscler. Thromb. Vasc. Biol. **17**: 1623-1629.

Aktories, K., U. Braun, S. Rosener, I. Just and A. Hall (1989). "The rho gene product expressed in *E. coli* is a substrate of botulinum ADP-ribosyltransferase C3." Biochem. Biophys. Res. Commun. **158**: 209-213.

Aktories, K. and J. Frevert (1987). "ADP-ribosylation of a 21-24 kDa eukaryotic protein(s) by C3, a novel botulinum ADP-ribosyltransferase, is regulated by guanine nucleotide." Biochem. J. **247**: 363-368.

Aktories, K., S. Rosener, U. Blaschke, S. Gursharan and G. S. Chhatwal (1988). "Botulinum ADP-ribosyltransferase C3 -Purification of the enzyme and characterization of the ADP-ribosylation reaction in platelet membranes." Eur. J. Biochem. **172**: 445-450.

Aktories, K., U. Weller and G. S. Chhatwal (1987). "Clostridium botulinum type C produces a novel ADP-ribosyltransferase distinct from botulinum C2 toxin." FEBS Lett. **212**: 109-113.

Allured, V. S., R. J. Collier, S. F. Carroll and D. B. McKay (1986). "Structure of exotoxin A of Pseudomonas aeruginosa at 3.0-Angstrom resolution." Proc. Natl. Acad. Sci. U. S. A. **83**: 1320-1324.

Barati, S., F. Chegini, P. Hurtado and R. A. Rush (2002). "Hybrid tetanus toxin C fragment-diphtheria toxin translocation domain allows specific gene transfer into PC12 cells." Exp. Neurol. **177**: 75-87.

Barret, T., B. Xiao, E. J. Dodson, G. Dodson, S. B. Ludbrook, K. Nurmahomed, S. J. Gamblin, A. Musacchio, S. Smerdon, J. and J. F. Eccleston (1992). "The structure of the GTPase-activating domain from p50rhoGAP." Nature **385**: 458-461.

Barth, H., J. C. Preiss, F. Hoffmann and K. Aktories (1998). "Characterization of the catalytic site of the ADP-ribosyltransferase *Clostridium botulinum* C2 toxin by site-directed mutagenesis." J. Biol. Chem. **273**: 29606-29511.

Beeckmans, S. (1999). "Chromatographic methods to study protein-protein interactions." Methods **19**: 278-305.

Bell, C. E. and D. Eisenberg (1996). "Crystal structure of diphtheria toxin bound to nicotinamide adenine dinucleotide." Biochemistry **35**: 1137-1149.

Bell, C. E. and D. Eisenberg (1997). "Crystal structure of nucleotide-free diphtheria toxin." Biochemistry **36**: 481-488.

Blasi, J., E. R. Chapman, E. Link, T. Binz, S. Yamasaki, P. De Camilli, T. C. Sudhof, H. Niemann and R. Jahn (1993). "Botulinum neurotoxin A selectively cleaves the synaptic protein SNAP-25." Nature **365**: 160-163.

Blaustein, R. O., W. J. Germann, A. Finkelstein and B. R. DasGupta (1987). "The N-terminal half of the heavy chain of botulinum type A neurotoxin forms channels in planar phospholipid bilayers." FEBS Lett. **226**: 115-120.

Blundell, T. L. and L. N. Johnson (1976). Protein Crystallography. London, Academic Press Inc. Ltd.

Bohmer, J., M. Jung, P. Sehr, G. Fritz, M. Popoff, I. Just and K. Aktories (1996). "Active site mutation of the C3-like ADP-ribosyltransferase from Clostridium limosum - analysis of glutamic acid 174." Biochemistry **35**: 282-289.

Bornhorst, J. A. and J. J. Falke (2000). "Purification of proteins using polyhistidine affinity tags." Methods Enzymol. **326**: 245-254.

Braun, U., B. Habermann, I. Just, K. Aktories and J. Vandekerckhove (1989). "Purification of the 22 kDa protein substrate of botulinum ADP-ribosyltransferase C3 from porcine brain cytosol and its characterization as a GTP-binding protein highly homologous to the *rho* gene product." FEBS Lett. **243**: 70-76.

Brünger, A. T. (1992). "Free *R* value: a novel statistical quantity for assessing the accuracy of crystal structures." Nature **355**: 472-475.

Brünger, A. T., P. D. Adams, G. M. Clore, W. L. DeLano, P. Gros, R. W. GrosseKunstleve, J. S. Jiang, J. Kuszewski, M. Nilges, N. S. Pannu, R. J. Read, L. M. Rice, T. Simonson and G. L. Warren (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr. **D54**: 905-921.

Brünger, A. T. and L. M. Rice (1997). "Crystallographic refinement by simulated annealing: Methods and applications." Methods Enzymol. **B 277**: 243-269.

Buchwald, B., A. Friebel, J. E. Galan, W. Hardt, A. Wittinghofer and K. Scheffzek (2002). "Structural basis for the reversible activation of a Rho protein by the bacterial toxin SopE." EMBO J. **21**: 3286-3295.

CCP4 (1994). "The CCP4 suite : Programs for protein crystallography." Acta Crystallogr. **D50**: 760-763.

Chardin, P., P. Boquet, P. Madaule, M. R. Popoff, E. J. Rubin and D. M. Gill (1989). "The mammalian G protein rhoC is ADP-ribosylated by *Clostridium botulinum* exoenzyme C3 and affects actin microfilaments in Vero cells." EMBO J. **8**: 1087-1092.

Chardin, P. and A. Tavitian (1986). "The *ral* gene: a new *ras* related gene isolated by the use of a synthetic probe." EMBO J. **5**: 2203-2208.

Chavan, A. J., Y. Nemoto, S. Narumiya, S. Kozaki and B. E. Haley (1992). "NAD^+ binding site of *Clostridium botulinum* C3 ADP-ribosyltransferase." J. Biol. Chem. **267**: 14866-14870.

Choe, S., M. J. Bennett, G. Fujii, P. M. Curmi, K. A. Kantardjieff, R. J. Collier and D. Eisenberg (1992). "The crystal structure of diphtheria toxin." Nature **357**: 216-222.

Coen, L., R. Osta, M. Maury and P. Brulet (1997). "Construction of hybrid proteins that migrate retrogradely and transynaptically into the central nervous system." Proc. Natl. Acad. Sci. U .S .A. **94**: 9400-9405.

Czech, A., T. Yamaguchi, L. Bader, S. Linder, K. Kaminski, M. Sugai and M. Aepfelbacher (2001). "Prevalence of Rho-inactivating epidermal cell differentiation inhibitor toxins in clinical *Staphylococcus aureus* isolates." J. Infect. Dis. **184**: 785-788.

Dauter, Z. (1997). "Data collection strategy." Methods Enzymol. **276**: 326-344.

Didsbury, J., R. F. Weber, G. M. Bokoch, T. Evans and R. Snyderman (1989). "*rac*, a novel *ras*-related family of proteins that are botulinum toxin substrates." J. Biol. Chem. **264**: 16378-16382.

Dobrenis, K., A. Joseph and M. C. Rattazzi (1992). "Neuronal lysosomal enzyme replacement using fragment C of tetanus toxin." Proc. Natl. Acad. Sci. U .S .A. **89**: 2297-2301.

Dolan, K. M., G. Lindenmayer and J. C. Olson (2000). "Functional comparison of the NAD binding cleft of the ADP-ribosylating toxins." Biochemistry **39**: 8266-8275.

Dolly, J. O., J. Black, R. S. Williams and J. Melling (1984). "Acceptors for botulinum neurotoxin reside on motor nerve terminals and mediate its internalization." Nature **307**: 457-460.

Duggan, M. J., C. P. Quinn, J. A. Chaddock, J. R. Purkiss, F. C. G. Alexander, S. Doward, S. J. Fooks, L. M. Friis, Y. H. J. Hall, E. R. Kirby, N. Leeds, H. J. Moulsdale, A.

Dickenson, G. M. Grenn, W. Rahman, R. Suzuki, C. C. Shone and K. A. Foster (2002). "Inhibition of release of neurotransmitters from rat dorsal root ganglia by a novel conjugate of a *Clostridium botulinum* toxin A endopeptidase fragment and *Erythina cristagalli* lectin." J. Biol. Chem. **277**: 34836-34852.

Eleopra, R., V. Tugnoli, O. Rossetto, D. De Grandis and C. Montecucco (1998). "Different time courses of recovery after poisoning with botulinum neurotoxin serotypes A and E in humans." Neurosci. Lett. **256**: 135-138.

Emsley, P., C. Fotinou, I. Black, N. F. Fairweather, I. G. Charles, C. Watts, E. Hewitt and N. W. Isaacs (2000). "The structures of the H(C) fragment of tetanus toxin with carbohydrate subunit complexes provide insight into ganglioside binding." J. Biol. Chem. **275**: 8889-8894.

Evans, H. R., J. M. Sutton, D. E. Holloway, J. A. Ayriss, C. C. Shone and K. R. Acharya (2003). "The structure of C3stau2 and its complex with NAD." J. Biol. Chem.: e-publication ahead of print.

Fenn, T. D., D. Ringe and G. A. Petsko (2003). "POVScript+: a program for model and data visualization using persistence of vision ray-tracing." J. Appl. Crystallogr. **36**: 944-947.

Fidyk, N. J. and R. A. Cerione (2002). "Understanding the catalytic mechanism of GTPase-activating proteins: Demonstration of the importance of the switch domain stabilization in the stimulation of GTP hydrolysis." Biochemistry **41**: 15644-15653.

Figueiredo, D. M., R. A. Hallewell, L. L. Chen, N. F. Fairweather, G. Dougan, J. M. Savitt and P. S. Fishman (1997). "Delivery of recombinant tetanus-superoxide dismutase proteins to the central nervous system neurons by retrograde axonal transport." Exp. Neurol. **145**: 546-554.

Fitzgerald, P. M. D. and N. B. Madsen (1986). "Improvement of limit of diffraction and useful x-ray lifetime of crystals of glycogen debranching enzyme." J. Cryst. Growth **76**: 600-606.

Fontinou, C., P. Emsley, I. Black, H. Ando, H. Ishida, M. Kiso, K. A. Sinha, N. F. Fairweather and N. W. Isaacs (2001). "The crystal structure of tetanus toxin Hc fragment complexed with a synthetic GT1b analogue suggests cross-linking between ganglioside receptors and the toxin." J. Biol. Chem. **276**: 32274-32281.

Foran, P. G., N. Mohammed, G. O. Lisk, S. Nagwaney, G. W. Lawrence, L. A. Smith, K. R. Aoki and J. O. Dolly (2003). "Evaluation of the therapeutic usefulness of botulinum neurotoxin B, C1, E & F compared to the long-lasting type A: Basis for

distinct durations of inhibition of exocytosis in central neurones." J. Biol. Chem. **278**: 1363-1371.

Foster, R., K.-Q. Hu, Y. Lu, K. M. Nolan, J. Thissen and J. Settleman (1996). "Identification of a novel human Rho protein with unusual properties: GTPases deficiency and in vivo farnesylation." Mol. Cell. Biol. **16**: 2689-2699.

Francis, J. W., R. H. J. Brown, D. Figueiredo, M. P. Remington, O. Castillo, M. A. Schwarzschild, P. S. Fishman, J. R. Murphy and J. C. vanderSpek (2000). "Enhancement of diphtheria toxin potency by replacement of the receptor binding domain with tetanus toxin C-fragment: a potential vector for delivering heterologous proteins to neurons." J. Neurochem. **74**: 2528-2536.

Genth, H., R. Gerhard, A. Maeda, M. Amano, K. Kaibuchi and K. Aktories (2003). "Entrapment of Rho ADP-ribosylated by *Clostridium botulinum* C3 exoenzyme in the Rho-GDI-1 complex." J. Biol. Chem. **278**: 28523-28527.

Goodnough, M. C., G. Oyler, E. A. Johnson, E. A. Neale, J. E. Keller, W. H. Tepp, M. Clark, S. Hartz and M. Adler (2002). "Development of a delivery vehicle for intracellular transport of botulinum neurotoxin antagonists." FEBS Lett. **25757**: 1-6.

Graham, D. L., P. N. Lowe, G. W. Grime, M. Marsh, K. Rittinger, S. Smerdon, J., S. J. Gamblin and J. F. Eccleston (2002). "MgF_3^- as a transition state analog of phosphoryl transfer." Chem. Biol. **9**: 375-361.

Green, D. W., V. M. Ingram and M. F. Perutz (1954). "The structure of haemoglobin IV. Sign determination by the isomorphous replacement method." Proc. Roy. Soc. London **A225**: 287-307.

Grizot, S., J. Faure, F. Fieschi, P. V. Vignais, M.-C. Dagher and E. Pebay-Peyroula (2001). "Crystal structure of the Rac1-rhoGDI complex involved in NADPH oxidase activation." Biochemistry **40**: 10007-10013.

Habermann, B., C. Mohr, I. Just and K. Aktories (1991). "ADP-ribosylation and de-ADP-ribosylation of the *rho* protein by *Clostridium botulinum* exoenzyme C3. Regulation by EDTA, guanine nucleotides and pH." Biochem. Biophys. Acta **1077**: 253-258.

Han, S., A. S. Arvai, S. B. Clancy and J. A. Tainer (2001). "Crystal structure and novel recognition motif of rho ADP-ribosylating C3 exoenzyme from Clostridium botulinum: structural insights for recognition specificity and catalysis." J. Mol. Biol. **305**: 95-107.

Han, S., J. A. Craig, C. D. Putnam, N. B. Carozzi and J. A. Tainer (1999). "Evolution and mechanism from structures of an ADP-ribosylating toxin and NAD complex." Nat. Struct. Biol. **6**: 932-936.

Han, S. and J. A. Tainer (2002). "The ARTT motif and a unified structural understanding of substrate recognition in ADP-ribosylating bacterial toxins and eukaryotic ADP-ribosyltransferases." Int. J. Med. Microbiol. **291**: 523-529.

Hanson, M. A. and R. C. Stevens (2000). "Cocrystal structure of synaptobrevin-II bound to botulinum neurotoxin type B at 2.0 A resolution." Nat. Struct. Biol. **7**: 687-692.

Helliwell, J. R. (1997). "Overview of synchrotron radiation and macromolecular crystallography." Methods Enzymol. **276**: 203-217.

Hendrickson, W. A., J. L. Smith and S. Sheriff (1985). "Direct phase determination based on anomalous scattering." Methods Enzymol. **115**: 41-55.

Henry, N. F. M. and K. Lonsdale (1952). International tables for X-ray crystallography Volume 1, The Kynoch Press.

Hoch, D. H., M. RomeroMira, B. E. Ehrlich, A. Finkelstein, B. R. DasGupta and L. L. Simpson (1985). "Channels formed by botulinum, tetanus, and diphtheria toxins in planar lipid bilayers: relevance to translocation of proteins across membranes." Proc. Natl. Acad. Sci. U. S. A. **82**: 1692-1696.

Houser, M. K., G. L. Sheean and A. J. Lees (1998). "Further studies using higher doses of botulinum toxin type F for torticollis resistant to botulinum toxin type A." J. Neurol. Neurosurg. Psychiatry **64**: 577-580.

Ihara, K., S. Muraguchi, M. Kato, T. Shimizu, M. Shirakawa, S. Kuroda, K. Kaibuchi and T. Hakoshima (1998). "Crystal structure of human RhoA in a dominantly active form complexed with a GTP analogue." J. Biol. Chem. **273**: 9656-9666.

Inoue, S., M. Sugai, Y. Murooka, S. Y. Paik, Y. M. Hong, H. Ohgai and H. Suginaka (1991). "Molecular cloning and sequencing of the epidermal cell differentiation inhibitor gene from *Staphylococcus aureus*." Biochem. Biophys. Res. Commun. **174**: 459-464.

Johnson, E. A. (1999). "Clostridial toxins as therapeutic agents: Benefits of nature's most toxic proteins." Annu. Rev. Microbiol. **53**: 551-575.

Jones, T. A. and M. Kjeldgaard (1997). "Electron-density map interpretation." Methods Enzymol. **277**: 173-208.

Jones, T. A., J. Y. Zou, S. W. Cowan and M. Kjeldgaard (1991). "Improved methods for building models in electron density maps and the location of errors in these models." Acta Crystallogr. **A47**: 110-119.

Jullien-Flores, V., Y. Mahe, G. Mirey, C. Leprince, B. Meunier-Bisceuil, A. Sorkin and J. H. Camonis (2000). "RLIP, an effector of the GTPase Ral, interacts with the AP2 complex: involvement of the Ral pathway in receptor endocytosis." J. Cell Sci. **113**: 2837-2844.

Jung, M., I. Just, J. van Damme, J. Vandekerckhove and K. Aktories (1993). "NAD-binding site of the C3-like ADP-ribosyltransferase from *Clostridium limosum*." J. Biol. Chem. **268**: 23215-23218.

Just, I., C. Mohr, G. Schallehn, L. Menard, J. R. Didsbury, J. Vandekerckhove, J. van Damme and K. Aktories (1992). "Purification and characterization of an ADP-ribosyltransferase produced by *Clostridium limosum*." J. Biol. Chem. **267**: 10274-10280.

Kandel, J., R. J. Collier and D. W. Chung (1974). "Interaction of Fragment A from diphtheria toxin with nicotinamide adenine dinucleotide." J. Biol. Chem. **249**: 2088-2098.

Keep, N. H., M. Barnes, I. NBarsukov, R. Badii, L. Lian, A. W. Segal, P. C. E. Moody and G. C. K. Roberts (1997). "A modulator of rho family G proteins, rhoGDI, binds these G proteins via and immunoglobulin -like domain and a flexible N-terminal arm." Structure **5**: 623-633.

Kikuchi, A., K. Yamamoto, T. Fujita and Y. Takai (1988). "ADP-ribosylation of the bovine brain *rho* by botulinum toxin type C1." J. Biol. Chem. **263**: 16303-16308.

Klein, A. W. (2003). "Complications, adverse reactions, and insights with the use of Botulinum toxin." Dermatol. Surg. **29**: 549-556.

Kleywegt, G. J. and T. A. Jones (1996). "*xdlMAPMAN* and *xdlDATAMAN* -Programs for reformatting, analysis and manipulation of biomacromolecular electron-density maps and reflection data sets." Acta Crystallogr. **D52**: 826-828.

Kleywegt, G. J. and T. A. Jones (1997). "Model building and refinement practice." Methods Enzymol. **B 277**: 208-230.

Koriazova, L. K. and M. Montal (2002). "Translocation of botulinum neurotoxin light chain protease through the heavy chain channel." Nat. Struct. Biol.

Kozaki, S., Y. Kamata, S. Watarai, T. Nishiki and S. Mochida (1998). "Ganglioside GT1b as a complementary receptor component for *Clostridium botulinum* neurotoxins." Micro. Pathogen. **25**: 91-99.

Kraulis, P. J. (1991). "MOLSCRIPT - A program to produce both detailed & schematic plots of protein structures." J. Appl. Crystallogr. **24**: 946-950.

Krueger, K. M. and J. T. Barbieri (1995). "The family of bacterial ADP-ribosylating exotoxins." Clin. Microbiol. Rev. **8**: 34-47.

Lacy, D. B., W. Tepp, A. C. Cohen, B. R. DasGupta and R. C. Stevens (1998). "Crystal structure of botulinum neurotoxin type A and implications for toxicity." Nat. Struct. Biol. **5**: 898-902.

Lamzin, V. S. and K. S. Wilson (1993). "Automated refinement of protein models." Acta Crystallogr. **D49**: 129-147.

Laskowski, R. A., M. W. MacArthur, D. M. Moss and J. M. Thornton (1993). "PROCHECK: a program to check the stereochemical quality of protein structures." J. Appl. Crystallog. **26**: 283-291.

Leatherbarrow, R. J. (2001). *GraFit*. Horley, Erithacus Software Ltd.

Lehmann, M., A. Fournier, I. Selles-Navarro, P. Dergham, A. Sebok, N. Leclerc, G. Tigyi and L. McKerracher (1999). "Inactivation of Rho signalling pathway promotes CNS axon regeneration." J. Neurosci. **19**: 7537-7547.

Leslie, A. G. W. (2001). Integration of macromolecular diffraction data. International tables for crystallography. M. G. Rossman and E. Arnold, Kluwer academic publishers. **Volume F: Crystallography of biological macromolecules**.

Longenecker, K. L., P. Read, U. Derewenda, Z. Dauter, X. Liu, S. Garrard, L. Walker, A. V. Somlyo, R. K. Nakamoto, A. P. Somlyo and Z. S. Derewenda (1999). "How RhoGDI binds Rho." Acta Crystallogr. **D55**: 1503-1515.

Malcom, K., C. Elliott and J. Exton (1996). "Evidence for Rho-mediated agonist stimulation of phospholipase D in Rat1 fibroblasts." J. Biol. Chem. **271**: 13135-13139.

Marvaud, J.-C., B. Stiles, A. Chenal, D. Gillet, M. Gibert, L. A. Smith and M. Popoff (2002). "*Clostridium perfringens* Iota toxin. Mapping of the Ia domain involved in docking with Ib and cellular internalization." J. Biol. Chem. **277**: 43659-43666.

Matthews, B. W. (1968). "Solvent content of protein crystals." J. Mol. Biol. **33**: 479-491.

McDonald, I. K. and J. M. Thornton (1994). "Satisfying hydrogen bonding potential in proteins." J. Mol. Biol. **238**: 777-793.

McPherson, A. (1985). "Crystallization of macromolecules: General principles." Methods Enzymol. **115**: 112-120.

McPherson, A. (1990). "Current approaches to macromolecular crystallization." Eur. J. Biochem. **189**: 1-23.

McPherson, A., S. Koszelak, H. Axelrod, J. Day, L. Robinson, M. McGrath, R. S. Williams and D. Cascio (1986a). "The effects of neutral detergents on the crystallization of soluble proteins." J. Cryst. Growth **76**: 547-553.

McPherson, A., S. Koszelak, H. Axelrod, J. Day, R. Williams, L. Robinson, M. McGrath and D. Cascio (1986b). "An experiment regarding crystallization of soluble proteins in the presence of β-octyl glucoside." J. Biol. Chem. **261**: 1969-1975.

Meacci, E., V. Vasta, J. Moorman, D. Bobak, P. Bruni, J. Moss and M. Vaughan (1999). "Effect of Rho and ADP-ribosylation factor GTPases on phospholipase D activity in intact human adenocarcinoma A549 cells." J. Biol. Chem. **274**: 18605-18612.

Menetrey, J., G. Flatau, E. A. Stura, J. B. Charbonnier, F. Gas, J. M. Teulon, M. H. Le Du, P. Boquet and A. Menez (2002). "NAD binding induces conformational changes in Rho ADP-ribosylating *Clostridium botulinum* C3 exoenzyme." J. Biol. Chem. **277**: 30960-30957.

Michaely, P., C. Mineo, Y. Ying and R. Andersn (1999). "Polarized distribution of endogenous Rac1 and RhoA at the cell surface." J. Biol. Chem. **274**: 21430-21436.

Misra, V. P. (2002). "The changed image of botulinum toxin." Brit. Med. J. **325**: 1188.

Montecucco, C. (1986). "How do tetanus and botulinum toxins bind to neuronal membranes?" Trends Biochem. Sci. **11**: 314-317.

Morii, N. and S. Narumiya (1995). "Preparation of native and recombinant *Clostridium botulinum* C3 ADP- ribosyltransferase and identification of Rho proteins by ADP-ribosylation." Methods Enzymol. **256**: 196-206.

Moskalenko, S., D. O. Henry, C. Rosse, G. Mirey, J. H. Camonis and M. A. White (2002). "The exocyst is a Ral effector complex." Nat. Cell Biol. **4**: 66-72.

Mott, H. R., D. Nietlispach, L. J. Hopkins, G. Mirey, J. H. Camonis and D. Owen (2003). "Structure of the GTPase-binding domain of Sec5 and elucidation of its Ral binding site." J. Biol. Chem. **278**: 17053-17059.

Mulligan, M. E., K. A. Murray-Leisure, B. S. Ribner, H. C. Standiford, H. F. John, J. A. Korvick, C. A. Kauffman and V. L. Yu (1993). "Methicillin-resistant *Staphylococcus aureus*: a consensus review of the microbiology, pathogenesis, and

epidemiology with implications for prevention and treatment." Am. J. Med. **94**: 313-231.

Munchau, A. and K. P. Bhatia (2000). "Uses of botulinum toxin injection in medicine today." Brit. Med. J. **320**: 161-165.

Murshudov, G. N., A. A. Vagin and E. J. Dodson (1997). "Refinement of Macromolecular Structures by the Maximum-Likelihood Method." Acta Crystallogr. **D53**: 240-255.

Nagahama, M., Y. Sakaguchi, S. Kobayashi, S. Ochi and J. Sakurai (2000). "Characterization of the enzymatic component of *Clostridium perfringens* ι–toxin." J. Bacteriol. **182**: 2096-2103.

Nakashima, S., K. Morinaka, S. Koyama, M. Ikeda, M. Kishida, K. Okawa, A. Iwamatsu, S. Kishida and A. Kikuchi (1999). "Small G protein Ral and its downstream molecules regulate endocytosis of EGF and insulin receptors." EMBO J. **18**: 3629-3642.

Narumiya, S., A. Sekine and M. Fujiwara (1988). "Substrate for botulinum ADP-ribosyltransferase, Gb, has an amino acid sequence homologous to a putative rho gene product." J. Biol. Chem. **263**: 17255-17257.

Nassar, N., G. R. Hoffman, D. Manor, J. C. Clardy and R. A. Cerione (1998). "Structures of Cdc42 bound to the active and catalytically compromised forms of Cdc42GAP." Nat. Struct. Biol. **5**: 1047-1052.

Navaza, J. (1994). "AMoRe: an automated package for molecular replacement." Acta Crystallogr. **A50**: 157-163.

Nemoto, Y., T. Namba, S. Kozaki and S. Narumiya (1991). "*Clostridium botulinum* C3 ADP-ribosyltransferase gene. Cloning, sequencing, and expression of a functional protein in *Escherichia coli*." J. Biol. Chem. **266**: 19312-19319.

Nishiki, T., Y. Kamata, Y. Nemoto, A. Omori, T. Ito, M. Takahashi and S. Kozaki (1994). "Identification of protein receptor for Clostridium botulinum type B neurotoxin in rat brain synaptosomes." J. Biol. Chem. **269**: 10498-10503.

Nishiki, T., Y. Tokuyama, Y. Kamata, Y. Nemoto, A. Yoshida, K. Sato, M. Sekiguchi, M. Takahashi and S. Kozaki (1996). "The high-affinity binding of Clostridium botulinum type B neurotoxin to synaptotagmin II associated with gangliosides GT1b/GD1a." FEBS Lett. **378**: 253-257.

Nobes, C. D., I. Lauritzen, M.-G. Mattei, S. Paris, A. Hall and P. Chardin (1998). "A new member of the Rho family, Rnd1, promotes disassembly of actin filament structures and loss of cell adhesion." J. Cell Biol. **141**: 187-197.

Notredame, C., D. Higgins and J. Heringa (2000). "T-Coffee: A novel method for multiple sequence alignments." J. Mol. Biol. **302**: 205-217.

Ohashi, Y. and S. Narumiya (1987). "ADP-ribosylation of a Mr 21,000 membrane protein by type D botulinum toxin." J. Biol. Chem. **262**: 1430-1433.

Ohta, Y., N. Suzuki, N. Nakamura, J. H. Hartwig and T. P. Stossel (1999). "The small GTPase RalA targets filimin to induce filopodia." Proc. Natl. Acad. Sci. U .S .A. **96**: 2122-2128.

Otwinowski, Z. and W. Minor (1997). "Processing of X-ray diffraction data collected in oscillation mode." Methods Enzymol. **276**: 307-326.

Paduch, M., F. Jelen and J. Otlewski (2001). "Structure of small G proteins and their regulators." Acta Biochim. Pol. **48**: 829-850.

Passador, L. and W. Iglewski (1994). "ADP-Ribosylating Toxins." Methods Enzymol. **235**: 617-631.

Paterson, H. F., A. J. Self, M. D. Garrett, I. Just, K. Aktories and A. Hall (1990). "Microinjection of recombinant p21rho induces rapid changes in cell morphology." J. Cell Biol. **111**: 1001-1007.

Persson, B. and P. Argos (1996). "Topology prediction of membrane proteins." Protein Sci. **5**: 363-71.

Popoff, M. R., D. Hauser, P. Boquet, M. W. Eklund and D. M. Gill (1991). "Characterization of the C3 gene of *Clostridium botulinum* types C and D and its expression in Escherichia coli." Infect. Immun. **59**: 3673-3679.

Potterton, E., P. Briggs, M. Turkenburg and E. J. Dodson (2003). "A graphical user interface to the CCP4 program suite." Acta Crystallogr. **D59**: 1131-1137.

Ramachandran, G. N., C. Ramakrishnan and V. Sasisekharan (1963). "Stereochemistry of polypeptide chain configurations." J. Mol. Biol. **7**: 95-99.

Resat, H., T. P. Straatsma, D. A. Dixon and J. H. Miller (2001). "The arginine finger of RasGAP helps Gln-61 aligns the nucleophilic water in GAP-stimulated hydrolysis of GTP." Proc. Natl. Acad. Sci. U .S .A. **98**: 6033-6038.

Rittinger, K., P. A. Walker, J. F. Eccleston, K. Nurmahomed, D. Owen, E. Laue, S. J. Gamblin and S. Smerdon, J. (1997a). "Crystal structure of a small G protein in complex with the GTPase-activating protein rhoGAP." Nature **388**: 693-697.

Rittinger, K., P. A. Walker, J. F. Eccleston, S. Smerdon, J. and S. J. Gamblin (1997b). "Structure at 1.65Å of RhoA and its GTPase-activating protein in complex with a transition-state analogue." Nature **389**: 758-762.

Rosener, S., G. S. Chhatwal and K. Aktories (1987). "Botulinum ADP-ribosyltransferase C3 but not botulinum neurotoxins C1 and D ADP-ribosylates low molecular mass GTP-binding proteins." FEBS Lett. **224**: 38-42.

Rossman, K. L., D. K. Worthylake, J. T. Snyder, D. P. Siderovski, S. L. Campbell and J. Sondek (2002). "A crystallographic view of interactions between Dbs and Cdc42: PH domain-assisted guanine nucleotide exchange." EMBO J. **21**: 1315-1326.

Rossmann, M. G. and D. M. Blow (1962). "The detection of sub-units within the crystallographic asymmetric unit." Acta Crystallogr. **15**: 24-31.

Rubin, E. J., D. M. Gill, P. Boquet and M. R. Popoff (1988). "Functional modification of a 21-kilodalton G protein when ADP- ribosylated by exoenzyme C3 of Clostridium botulinum." Mol. Cell. Biol. **8**: 418-26.

Sachdev, D. and J. M. Chirgwin (2000). "Fusions to maltose-binding protein: Control of folding and solubility in protein purification." Methods Enzymol. **326**: 312-321.

Saito, Y., Y. Nemoto, T. Ishizaki, N. Watanabe, N. Morii and S. Narumiya (1995). "Identification of Glu173 as the critical amino acid residue for the ADP-ribosyltransferase activity of *Clostridium botulinum* C3 exoenzyme." FEBS Lett. **371**: 105-109.

Sakurai, J., M. Nagahama, J. Hisatsune, N. Katunuma and H. Tsuge (2003). "*Clostridium perfringens* ι–toxin ADP-ribosyltransferase: structure and mechanism of action." Advan. Enzyme Regul. **43**: 361-377.

Scheffzek, K., I. Stephan, O. N. Jensen, D. Illenberger and P. Gierschik (2000). "The Rac-RhoGDI complex and the structural basis for the regulation of Rho proteins by RhoGDI." Nat. Struct. Biol. **7**: 122-126.

Scheuring, J., P. J. Berti and V. L. Schramm (1998). "Transition-state structure for the ADP-ribosylation of recombinant $G_{\alpha i1}$ subunits by Pertussis toxin." Biochemistry **37**: 2748-2758.

Scheuring, J. and V. L. Schramm (1997a). "Kinetic isotope effect characterization of the transition state for oxidized nicotinamide adenine dinucleotide hydrolysis by pertussis toxin." Biochemistry **36**: 4526-4534.

Scheuring, J. and V. L. Schramm (1997b). "Pertussis toxin: Transition state analysis for ADP-ribosylation of G-protein peptide $_{\alpha i3}$C20." Biochemistry **36**: 8215-8223.

Schiavo, G., A. Santucci, B. R. Dasgupta, P. P. Mehta, J. Jontes, F. Benfenati, M. C. Wilson and C. Montecucco (1993). "Botulinum neurotoxins serotypes A and E

cleave SNAP-25 at distinct COOH- terminal peptide bonds." FEBS Lett. **335**: 99-103.

Schneider, H., M. Groves, C. Muhle, P. N. Reynolds, A. Knight, M. Thermis, J. Carvajal, F. Scaravilli, D. T. Curiel, N. F. Fairweather and C. Coutelle (2000). "Retargeting of adenoviral vectors to neurons using the HC fragment of tetanus toxin." Gene Ther. **7**: 1584-1592.

Sehr, P., G. Joseph, H. Genth, I. Just, E. Pick and K. Aktories (1998). "Glucosylation and ADP ribosylation of rho proteins: effects on nucleotide binding, GTPase activity, and effector coupling." Biochemistry **37**: 5296-5304.

Sekine, A., M. Fujiwara and S. Narumiya (1989). "Asparagine residue in the rho gene product is the modification site for botulinum ADP-ribosyltransferase." J. Biol. Chem. **264**: 8602-8605.

Sheldrick, G. M. and R. O. Gould (1995). Acta Crystallogr. **B51**: 423-431.

Sheldrick, G. M. and T. R. Schneider (1997). "SHELXL: High-resolution refinement." Methods Enzymol. **277**: 310-343.

Shimizu, T., K. Ihara, R. Maesaki, S. Kuroda, K. Kaibuchi and T. Hakoshima (2000). "An open conformation of switch I revealed by the crystal structure of an Mg^{2+} -free form of RHOA complexed with GDP." J. Biol. Chem. **275**: 18311-18317.

Shone, C. C., P. Hambleton and J. Melling (1987). "A 50-kDa fragment from the NH2-terminus of the heavy subunit of Clostridium botulinum type A neurotoxin forms channels in lipid vesicles." Eur. J. Biochem. **167**: 175-180.

Shone, C. C. and H. S. Tranter (1995). "Growth of clostridia and preparation of their neurotoxins." Curr. Top. Microbiol. Immunol. **195**: 143-160.

Simpson, L. L. (2000). "Identification of the characteristics that underlie botulinum toxin potency: implications for designing novel drugs." Biochimie **82**: 943-953.

Simpson, L. L., J. A. Coffield and N. Bakry (1994). "Inhibition of vacuolar adenosine triphosphatase antagonizes the effects of clostridial neurotoxins but not phospholipase A2 neurotoxins." J. Pharmacol. Exp. Ther. **269**: 256-262.

Slama, J. T. and A. M. Simmons (1988). "Carbanicotinamide Adenine Dinucleotide: Synthesis and enzymological properties of a carbocyclic analogue of oxidised nicotinamide adenine dinucleotide." Biochemistry **27**: 183-193.

Smith, D. B. (2000). "Generating fusions to glutathione S-transferase for protein studies." Methods Enzymol. **326**: 254-270.

Snyder, J. T., D. K. Worthylake, K. L. Rossman, L. Betts, W. M. Pruitt, D. P. Siderovski, D. J. Der and J. Sondek (2002). "Structural basis for the selective activation of Rho GTPases by Dbl exchange factors." Nat. Struct. Biol. **9**: 468-475.

Stura, E. A. and I. A. Wilson (1991). "Applications of the streak seeding technique in protein crystallization." J. Cryst. Growth **110**: 270-282.

Sugai, M., T. Enomoto, K. Hashimoto, K. Matsumoto, Y. Matsuo, H. Ohgai, Y. M. Hong, S. Inoue, K. Yoshikawa and H. Suginaka (1990). "A novel epidermal cell differentiation inhibitor (EDIN): purification and characterization from *Staphylococcus aureus*." Biochem. Biophys. Res. Commun. **173**: 92-98.

Sugai, M., T. Enomoto, M. Miyake and H. Suginaka (1987). "Extracellular products of *Staphylococcus aureus* reversibly inhibit the terminal differentiation of cultured mouse epidermal cells." Cell Struct. Funct. **12**: 395-399.

Sugai, M., K. Hashimoto, A. Kikuchi, S. Inoue, H. Okumura, K. Matsumoto, Y. Goto, H. Ohgai, K. Moriishi, B. Syuto and et al. (1992). "Epidermal cell differentiation inhibitor ADP-ribosylates small GTP- binding proteins and induces hyperplasia of epidermis." J. Biol. Chem. **267**: 2600-2604.

Sugihara, K., S. Asano, K. Tanaka, A. Iwamatsu, K. Okawa and Y. Ohta (2002). "The exocyst complex binds the small GTPase RalA to mediate filopodia formation." Nat. Cell Biol. **4**: 73-78.

Swaminathan, S. and S. Eswaramoorthy (2000). "Structural analysis of the catalytic and binding sites of *Clostridium botulinum* neurotoxin B." Nat. Struct. Biol. **7**: 693-699.

Thaller, C., G. Eicjele, L. H. Weaver, E. Wilson, R. Karlsson and J. N. Jansonius (1985). "Seed enlargement and repeated seeding." Methods Enzymol. **115**: 132-135.

Tsuge, H., M. Nagahama, H. Nishimura, J. Hisatsune, Y. Sakaguchi, Y. Itogawa, N. Katunuma and J. Sakurai (2003). "Crystal structure and site directed mutagenesis of enzymatic components from *Clostridium perfringens* Iota-toxin." J. Mol. Biol. **325**: 471-483.

Turton, K., J. A. Chaddock and K. R. Acharya (2002). "Botulinum and tetanus neurotoxins: structure, function and therapeutic utility." Trends. Biochem. Sci. **27**: 552-558.

Vagin, A. and A. Teplyakov (1997). "MOLREP: an automated program for molecular replacement." J. Appl. Cryst. **30**: 1022-1025.

Watanabe, Y., M. Morimatsu and B. Syuto (2000). "The evaluation of the potential of botulinum C3 enzyme as an exogenous differentiation inducing factor to neurons." J. Vet. Med. Sci. **62**: 473-478.

Wei, Y., Y. Zhang, U. Derewenda, X. Liu, W. Minor, R. K. Nakamoto, A. V. Somlyo, A. P. Somlyo and Z. D. Derewenda (1997). "Crystal structure of RhoA-GDP and its structural implications." Nat. Struct. Biol. **4**: 699-703.

Wilde, C. and K. Aktories (2001). "The Rho-ADP-ribosylating C3 exoenzyme from *Clostridium botulinum* and related C3-like transferases." Toxicon **39**: 1647-1660.

Wilde, C., H. Barth, P. Sehr, L. Han, M. Schmidt, I. Just and K. Aktories (2002a). "Interaction of the Rho-ADP-ribosylating C3 exoenzyme with RalA." J. Biol. Chem. **277**: 14771-14776.

Wilde, C., G. S. Chhatwal and K. Aktories (2002b). "C3stau, a new member of the family of C3-like ADP-ribosyltransferases." Trends Microbiol. **10**: 5-7.

Wilde, C., G. S. Chhatwal, G. Schmalzing, K. Aktories and I. Just (2001). "A novel C3-like ADP-ribosyltransferase from *Staphylococcus aureus* modifying RhoE and Rnd3." J. Biol. Chem. **276**: 9537-9542.

Wilde, C., H. Genth, K. Aktories and I. Just (2000). "Recognition of RhoA by *Clostridium botulinum* C3 exoenzyme." J. Biol. Chem. **275**: 16478-16483.

Wilde, C., I. Just and K. Aktories (2002c). "Structure-function analysis of the Rho-ADP-ribosylating exoenzyme C3stau2 from *Staphylococcus aureus*." Biochemistry **41**: 1539-1544.

Wilde, C., M. Vogelsgesang and K. Aktories (2003). "Rho-specific *Bacillus cereus* ADP-ribosyltransferase C3cer cloning and characterization." Biochemistry **42**: 9693-9702.

Williamson, K. C., L. A. Smith, J. Moss and M. Vaughan (1990). "Guanine nucleotide-dependent ADP-ribosylation of soluble *rho* catalysed by *Clostridium botulinum* C3 ADP-ribosyltransferase." J. Biol. Chem. **265**: 20807-20813.

Winton, M. J., C. I. Dubreuil, D. Lasko, N. Leclerc and L. McKerracher (2002). "Characterization of new cell permeable C3-like proteins that inactivate Rho and stimulate neurite outgrowth on inhibitory substrates." J. Biol. Chem. **277**: 32820-32829.

Worthylake, D. K., K. L. Rossman and J. Sondek (2000). "Crystal structure of Rac1 in complex with the guanine nucleotide exchange region of Tiam1." Nature **408**: 682-688.

www.povray.org.

Yamada, Y., M. Sugai, M. Woo, N. Nishida and T. Sugimoto (2001). "Acquired subglottic stenosis caused by methicillin resistant *Staphylococcus aureus* that produce epidermal cell differentiation inhibitor." Arch. Dis. Child Fetal Neonatal Ed. **84**: F38-39.

Yamaguchi, T., T. Hayashi, H. Takami, M. Ohnishi, T. Murata, K. Nakayama, K. Asakawa, M. Ohara, H. Komatsuzawa and M. Sugai (2001). "Complete nucleotide sequence of a *Staphylococcus aureus* exfoliative toxin B plasmid and identification of a novel ADP-ribosyltransferase, EDIN-C." Infect. Immun. **69**: 7760-7771.

Yamaguchi, T., K. Nishifuji, M. Sasaki, Y. Fudaba, M. Aepfelbacher, T. Takata, M. Ohara, H. Komatsuzawa, M. Amagai and M. Sugai (2002). "Identification of the *Staphylococcus aureus* etd pathogenicity island which encodes a novel exfoliative toxin, ETD, and EDIN-B." Infect. Immun. **70**: 5835-5845.

Zhang, K. Y. J., K. Cowtan and P. Main (1997). "Combining constraints for electron-density modification." Methods Enzymol. **277**: 53-64.

Zhang, R.-G., D. L. Scott, M. L. Westbrook, S. Nance, B. D. Spangler, G. G. Shipley and E. M. Westbrook (1995). "The three-dimensional crystal structure of cholera toxin." J. Mol. Biol. **251**: 563-573.

Appendix
Crystallographic programs

HKL Package (Otwinowski and Minor 1997)

This package comprises three programs: *XdisplayF*, for visualisation of the diffraction images, *Denzo*, for indexing the data and assigning intensities to hkl reflections and *Scalepack*, for scaling the images to a uniform background and calculating statistics for data quality. These programs have now been combined in the gui controlled HKL2000 which is available at the SRS, Daresbury.

CCP4 (CCP4 1994)

The collaborative computational project 4 (CCP4) makes freely available publicly funded programs for all areas of macromolecular crystallography. The most well developed of these are also now available through the CCP4 gui (Potterton *et al.* 2003) which is designed for easy tracking of structure determination steps and data harvesting. The following list represents the programs I used, not a comprehensive list of the programs available.

Scalepack2mtz converts the scaled output of the HKL suite (sca file) into mtz format required for CCP4 programs.

Truncate calculates structure factor amplitudes from the observed intensities.

Mtz2various converts the mtz format reflection data to other formats such as fob for CNS.

Matthews calculates the Matthews coefficient (Matthews 1968) from the size of the unit cell, the unit cell symmetry and the molecular weight of the molecule. From this it estimates the solvent content of the crystal and the number of molecules per asymmetric unit.

AMoRE (Navaza 1994) is designed as a fast molecular replacement program. The program combines six sub-programs. *Sorting,* reformats the mtz reflection file for compatibility with the program, *Tabling*, uses an efficient algorithm for fast computation of the structure factors of the model, *Roting*, is the fast rotation function, *Traing,* the fast translation function, *Fitting*, a fast rigid body refinement of the rotation and translation

outputs and *pdbset*, orients the model using the calculated parameters. These programs are also available as autoAMoRE on the CCP4 gui.

Molrep (Vagin and Teplyakov 1997) Molrep, which can be accessed via the CCP4i gui, also performs rotation and translation functions. Additionally, it has a packing function which allows it to eliminate solutions with symmetry clashes.

Refmac (Murshudov *et al.* 1997) performs rigid body, restrained, unrestrained and tls refinement of either a least squares or maximum likelihood function. Addition of water molecules is achieved by an implementation of ARP/wARP (Lamzin and Wilson 1993), an **A**utomated **R**efinement **P**rocedure for tracing protein chains in high resolution experimental maps.

FFT is used to calculate the maps from Refmac refinement using the mtz and pdb output. Map type, size and format can be specified as appropriate.

Procheck (Laskowski *et al.* 1993) and SFcheck calculates a number of plots and statistics for data analysis and structural validation. SFcheck analyses the structure factors for twinning and anisotropic distribution. Procheck performs stereochemical analysis including Ramachandran plots (Ramachandran *et al.* 1963)

CNS, the Crystallography and NMR suite (Brünger *et al.* 1998) contains many programs for data manipulation, phase determination, density modification, map calculation and refinement. Refinement includes rigid body refinement, minimisation via conjugate gradient optimisation of a maximum likelihood target and simulated annealing (Brünger and Rice 1997). It also contains a program for producing suitable coordinate files for pdb submission.

SHELXL (Sheldrick and Schneider 1997) is a refinement program which does full matrix minimisation of a least squares target. It works best with high resolution data (>2Å).

Moleman manipulates coordinate files and performs a variety of analyses on them including B-factor averaging.

Mapman (Kleywegt and Jones 1996) normalises and converts formats of maps from various refinement programs.

O (Jones *et al.* 1991) is a graphical model building program where maps can be displayed and models built and manipulated. LSQ superpositioning can also be performed within the program.

HBPLUS (McDonald and Thornton 1994) calculates potential hydrogen bonds for structures and lists the distances and bond angles for further analysis.

Molscript (Kraulis 1991) creates pictures of macromolecules. The input file created describes the type of representation, colour and orientation of the molecules in the pdb file. This file, which can also specify labels and rendering, is used to create an image file.

Grasp is a graphics program which can display molecular surfaces and calculate parameters for them such as surface potential. These surfaces can be output for conversion into image files.

Povray (www.povray.org) is a freely distributed 3D graphics drawing and rendering program. It can be used, via POVScript+ (Fenn *et al.* 2003) to render image files from Molscript and Grasp.